UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA

IDENTIFICATION OF GENETIC RISK FACTORS FOR BEHÇET'S DISEASE

Joana Gonçalves de Gouveia Maia Xavier

Tese orientada por:

Professora Doutora Ana Sofia A. Oliveira

Instituto de Medicina Molecular/Faculdade de Medicina da Universidade de Lisboa

Doutoramento em Ciências Biomédicas

Especialidade de Genética

2013

As opiniões expressas no presente documento são da exclusiva responsabilidade do seu autor.

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 17 de Setembro de 2013.

ACKNOWLEDGEMENTS

Queria começar por agradecer à minha orientadora, Sofia Oliveira, por me ter dado a oportunidade de fazer o doutoramento no seu grupo, no Instituto de Medicina Molecular, e neste projecto, pelo qual me apaixonei durante o meu mestrado, e que no fundo e muito pessoalmente sempre senti que era meu. Sinto que fui privilegiada ao ter tido a oportunidade de trabalhar num projecto tão interessante e dinâmico ao nível de abordagens aplicadas e dos recursos disponíveis, e também ao ter tido a oportunidade de participar em inúmeros congressos nacionais e internacionais, e de frequentar cursos especializados, nos quais aprendi imenso. Também não posso deixar de salientar que o enorme gosto que tive em trabalhar neste projecto se deveu em muito à orientação da Sofia, que além de me ter dado todas as ferramentas necessárias para eu crescer cientificamente, me orientou sempre com base numa relação de confiança, e com um grande sentido ético, e me deu a oportunidade de pensar livremente e ser independente na minha investigação, que muito valorizei.

Queria também agradecer ao Prof. Coutinho, ao Dr. Jorge Crespo e à Dra. Francisca Fontes que tiveram a ideia e a vontade de iniciar este projecto, nesta doença tão pouco estudada e desconhecida para muitos, e sem o interesse e curiosidade científica dos quais este projecto nunca teria começado.

Queria deixar um grande agradecimento a todos os meus colegas de grupo passados e presentes, que me acompanharam durante esta jornada e que fizeram com que o laboratório fosse um sítio bom para estar: Tiago Krug, Benedita Fonseca, Sara Violante, Alexandra Rosa, Patrícia Abrantes, Madalena Martins, João Sobral, Nádia Rei, Inês Sousa, Vânia Francisco e Mafalda Matos - obrigada pelo companheirismo e amizade.

Obrigada a todos os meus amigos, especialmente à Vânia, Érica, Vera, Romina, Rita e Nádia, pelos cafés, jantaradas, saídas, praia e corridas, que contribuíram para que a minha vida nestes últimos quatro anos e meio não fosse só ciência - girls you rock!

Um grande obrigada aos meus pais e irmão que além de estarem sempre disponíveis para me alegrar (e facilitar) a vida, me apoiam em todas as decisões que tomo e confiam totalmente em mim - vocês são a melhor família do Mundo!

Last but not least, um obrigada especial ao Pedro que me acompanhou durante todo este projecto, celebrando os meus sucessos como se fossem dele e apoiando-me em todos os momentos - sem ti não seria igual.

“Somewhere, something incredible is waiting to be known” — Carl Sagan

GENERAL INDEX

ABSTRACT

Background: Behçet's disease (BD) is a complex disorder characterized by a generalized vasculitis, whose pathophysiology remains unclear. The identification of genes involved in BD can help to elucidate the disease mechanisms and, ultimately, result in diagnostic and treatment advances.

Objectives: To identify genetic risk factors implicated in BD susceptibility.

Methods: We performed four independent studies: 1) Analysis of the role of the mitochondrial genome by testing the association of mitochondrial haplogroups and variants with BD risk in 615 Iranian BD cases and 434 controls; 2) Follow-up of *IL10* and *IL23R-IL12RB2* associations, previously identified as BD risk factors, in 973 Iranian BD cases and 637 controls; 3) Gene expression profiling in 15 Portuguese BD cases and 14 controls and association testing of the differentially expressed genes in 976 Iranian BD cases and 839 controls; 4) A genome-wide association study for the Iranian population in DNA pools of 292 BD cases and 294 controls and replication of the association findings in 684 BD cases and 532 controls.

Results: We identified a novel association of BD with the mitochondrial 12S rRNA gene (7.00E-03<*P*<3.80E-02); replicated the association of *IL10* (*P*=2.53E-02) and *IL23R-IL12RB2* loci (1.93E-06<*P*<1.78E-05) and identified the region upstream *IL23R* as the most associated one; identified *EREG*, *AREG* and *NRG1* (members of the Neuregulin signalling) as down-regulated in BD patients, found a novel association in the *EREG-AREG* locus (*P*=2.51E-02) and replicated three associations at *NRG1* (6.61E-04<P_{meta}<2.10E-03*);* and identified five coding variants at *FUT2* associated with BD (2.97E-06<*P*<1.34E-04).

Conclusions: During the course of this project we have uncover the mitochondrial genome, the neuregulin signaling and the *FUT2* gene as novel players in BD susceptibility that may contribute to the abnormal immunological response observed in BD patients. We have further contributed to establish the *IL10* and *IL23R* loci as worldwide risk factors for Behçet's disease.

Key-words: Behçet's disease; genetic susceptibility, gene expression; association studies; etiopathogenesis.

RESUMO

Introdução: A doença de Behçet (DB) é uma doença complexa caracterizada por uma vasculite generalizada, cuja patofisiologia é ainda pouco conhecida. A identificação de genes envolvidos na DB pode ajudar a elucidar os mecanismos de doença levando a avanços a nível do diagnóstico e tratamento.

Objetivos: Identificar factores de risco genético para a DB.

Métodos: Realizou-se quatro estudos independentes: 1) Análise do papel do genoma mitocondrial, onde se testou a associação da DB com haplogrupos e variantes mitocondriais, em 615 casos e 434 controlos Iranianos; 2) *Follow-up* da associação dos genes *IL10* e *IL23R-IL12RB*, previamente identificados como factores de risco para a DB, em 973 casos e 637 controlos Iranianos; 3) Estudo de perfis génicos em 15 casos com DB e 14 controlos Portugueses e teste da associação dos genes diferencialmente expressos em 976 casos e 839 controlos Iranianos; 4) Estudo de associação no genoma inteiro (GWAS) para a DB na população Iraniana, em pools de ADN com 292 casos e 294 controlos, e replicação das associações identificadas em 684 casos e 532 controlos.

Resultados: Identificou-se uma nova associação da DB com o gene mitocondrial *12S rRNA* (7.00E-03<*P*<3.80E-02); Replicou-se a associação do *IL10* (*P*=2.53E-02) e *locus IL23R-IL12RB2* (1.93E-06>*P*<1.78E-05) e identificou-se a região regulatória do *IL23R* como a mais mais fortemente associada; Verificou-se a sub-expressão do *EREG*, *AREG* e *NRG1* (pertencentes à via da Neuregulina) em pacientes com DB, a associação do *locus EREG-AREG* (*P*=2.51E-02) e replicou-se três associações no *NRG1* (6.61E-04<P_{meta}<2.10E-03); Identificou-se cinco variantes codificantes no *FUT2* associadas com a DB (2.97E-06<*P*<1.34E-04).

Conclusões: Durante o curso deste projecto, identificou-se o genoma mitocondrial, a via da Neuregulina e o *FUT2* como novos intervenientes na suscetibilidade para a DB que podem contribuir para a resposta imunológica alterada observada nos pacientes com DB. Adicionalmente, contribui-se para estabelecer o *IL10* e *IL23R* como fatores de risco mundiais para a DB.

Palavras-chave: Doença de Behçet; susceptibilidade genética, estudos de associação; expressão génica; etiopatogénese.

INDEX OF FIGURES

Chapter 1 - General introduction

Chapter 3 - Association of mitochondrial polymorphism m.709G>A with Behçet's disease

Chapter 4 - Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients

Chapter 5 - Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility

Chapter 6 - FUT2: filling the gap between genes and environment in Behçet's disease?

Chapter 7 - General discussion

INDEX OF TABLES

Chapter 1 - General introduction

Chapter 3 - Association of mitochondrial polymorphism m.709G>A with Behçet's disease

Chapter 4 - Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients

Chapter 5 - Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility

Chapter 6 - *FUT2*: filling the gap between genes and environment in Behçet's disease?

LIST OF ABBREVIATIONS

AECE	Endothelial cells antibodies
AIMs	Ancestry informative markers
ANOVA	Analysis of variance
APC	Antigen presenting cell
AU	Anterior uveitis
AREG	Amphiregulin
BD	Behçet's disease
CARD15	Caspase recruitment domain-containing protein 15
CI	Confidence interval
CNS	Central nervous system
CNV	Copy number variant
CPLX1	Complexin-1
CPVL	Carboxypeptidase, vitellogenic-like
CSF	Cerebrospinal fluid
CTLA4	Cytotoxic T-lymphocyte antigen 4
d.f.	Degee of freedom
EN	Erithema modosum
eNOS	Endothelial nitric oxide synthase
EREG	Epiregulin
FC	Fold-change
FCRL3	Fc receptor-like protein 3
FDR	False discovery rate
FCRL3	Fc receptor-like protein 3
FUT2	Fucosyltransferase 2
GA	Genital aphthosis
GIMAP4	GTPase, IMAP family member 4
GIMAP7	GTPase, IMAP family member 7
GST	Glutathione S-transferase
GWAS	Genome-wide association study
HC	Hierarchical clustering
HLA	Human leukocyte antigen
htSNPs	Haplotype tagging SNPs

HSP	Heat chock protein
HWE	Hardy Weinberg equilibrium
IBD	Identical by descent
ICAM1	Intercellular adhesion molecule-1
ICBD	International Criteria for Behçet's disease
IFNγ	Interferon gamma
IL	Interleukin
IL23R	Interleukin-23 Receptor
IL12RB2	Iinterleukin 12 receptor, beta 2 subunit
ISG	International Study Group
Kb	Kilobases
KIAA1529	Chromosome 9 open reading frame 174
KIRs	Killer cell immunoglobulin-like receptors
KLRK1	Killer cell lectin-like receptor subfamily K, member 1
LD	Linkage disequilibrium
LOC100129342	Hypothetical protein LOC100129342
LOD	Log of Odds
MAF	Minor allele frequency
MDR	Multifactor dimensionality reduction
MHC	Major histocompatibility complex
MICA	MHC class I chain-related gene A
mRNA	Messenger RNA
MS	Microsatellite
NGS	Next generation sequencing
NRG1	Neuregulin-1
NK	Natural killer cells
OR	Odds ratio
PB	Peripheral blood
PBMCs	Peripheral blood mononuclear cells
PCA	Principal component analysis
PCR	Polymerase chain reaction
PF	Pseudofolliculitis
PLCB1	Phospholipase C, beta 1
PP	Pathergy phenomenon

PSORS1C1	Psoriasis susceptibility 1 candidate 1
PU	Posterior uveitis
RAS	Relative allele score
RNA	Ribonucleic acid
ROA	Recurrent oral aphthosis
ROS	Reactive oxygen species
RR	Relative risk
RV	Retinal vasculitis
SLC11A1	Solute carrier family 11 member 1
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
STAT4	Signal transducer and activator of transcription 4
Th1	T helper cells type-1
Th2	T helper cells type-2
Th17	T helper cells type-17
TLR4	Toll-like receptor 4
TNF	Tumor necrosis factor, gene
TNFα	Tumor necrosis factor alpha
TNFβ	Tumor necrosis factor beta
Tregs	Regulatory T cells
UBAC2	Ubiquitin-associated domain containing, 2
UBASH2B	Ubiquitin-associated and SH3 domain containing, B
VEGF	Vascular endothelial growth factor
VMs	Vascular manifestations
WWC1	WW domain-containing protein 1
YTHDC1	YTH domain-containing protein 1 gene

STATEMENT OF WORK

The work described on this thesis was designed, performed and analyzed by the candidate with contributions from collaborators that should be addressed.

The candidate participated in the sample preparation, genotyping, statistical analyses and discussion of the results, as well as writing of the published and unpublished articles reported in this thesis, where she contributed as first author. The Iranian individuals participating in this study were recruited at the Shariati Hospital in Tehran, Iran where the clinical assessment, collection of blood samples and extraction of DNA were performed by the Iranian collaborators, where the special contribution of Dr. Bahar Sadeghi Abdollahi and Dr. Fereydoun Davatchi should be adressed. The RNA from the Portuguese individuals participating in this study was collected at several Portuguese Hospitals and Institutes such as the Instituto Português de Reumatologia, Hospitais da Universidade de Coimbra, Hospital Infante D. Pedro and Instituto Português do Sangue, with the collaboration of medical doctors, specially Dr. Jorge Crespo and Dr. José Vaz-Patto, and of members of Sofia Oliveira's lab at Instituto Gulbenkian de Ciência and Instituto de Medicina Molecular, mainly Tiago Krug and Benedita Fonseca. I would like also to emphasize the special contribution of Alexandra Rosa for the design and analysis of the mitochondrial genome experiments reported in Chapter 3, Tiago Krug for the experimental work and analysis of the gene expression study reported in Chapter 5 and João Sobral, Mafalda Matos and Inês Sousa for the experimental work of the genome-wide association study reported in Chapter 6. Finally I would like to address the fundamental contribution of my supervisor, Prof. Sofia Oliveira, for the conception and design of the experiments, discussion and interpretation of data, and critical revising of all the work reported in this thesis.

THESIS OUTLINE

The work presented in this thesis was carried out between January 2009 and December 2012 at the Instituto de Medicina Molecular under supervision of Prof. Sofia A. Oliveira.

This thesis is organized in seven chapters, preceded by a summary in Portuguese and an abstract in English briefly describing the work developed. Chapter 1 consists of a general introduction and is divided in three main sections: a state of the art about Behçet's disease, with special focus on the immune and genetics aspects involved in the disease aetiology; an introduction to the genomic approaches used in the study of complex diseases; and a last section with the references cited along the chapter. In Chapter 2 are described the specific aims of the thesis. From Chapter 3 to 6 are presented the original results obtained during the course of this thesis. Chapter 3, 4 and 5 consists of data published in international scientific journals, were the candidate was first author, and therefore the main manuscript is shown as published in the first section of each chapter and the supplementary material is shown in the second section. Chapter 6 consists of unpublished data and is organized in seven main sections: abstract, introduction, patients and methods, results, discussion, references and supplementary material. Chapter 7 comprises a generalized discussion about the experimental design, and results obtained during the course of this thesis, and is divided in five main sections: main findings, limitations of the experimental design, follow-up studies, final remarks, and references.

CHAPTER 1

General Introduction

1.1 BEHÇET'S DISEASE

1.1.1 Historic perspective

Behçet's disease (BD) was first described in 1930 by a Greek ophthalmologist – Benediktos Adamantiades, that presented in a meeting a lecture entitled "A case of relapsing iritis with hypopyon" and brought together the oral and genital ulcers, the arthritis and the ocular signs as signs of a single disease [Adamantiades, 1930; reviewed in Zouboulis, 2002]. In 1937, Hulusi Behçet, a Turkish dermatologist, described the symptomatic complex of hypopyon, iritis and orogenital aphthosis, and postulated a possible viral aetiology. He insisted on the clinical individuality of the disease, earning the merit of having the disease named after him [Behçet, 1937; Behçet *et al*., 2010]. However, the first description may come from 2500 years earlier, when Hippocrates of Kos (470–377 BC) described in his *Epidemion*, book 3, case 7, a disease characterized by aphthous ulcers that would also cause genital and ophthalmological problems and which was endemic in Asia minor [Feigenbaum, 1956].

1.1.2 Epidemiology

The geographical distribution of BD is distinctive: it is most prevalent along the historical Silk Road, an ancient trading route between the Mediterranean (Portugal and Spain) and East Asia (China) (Figure 1), suggesting perhaps that an as-yet-unknown genetically determined factor was spread via the migration of old nomadic tribes [Al-Otaibi *et al*., 2005]. Iran is situated in the middle of the Silk Road and was the carrefour between East and West. Nowadays, due to immigration, BD can be seen in any country of the world [Davatchi *et al*., 2010a].

In Turkey, the country with the highest prevalence of the disease, BD is estimated to have a prevalence of 110-420 per 100,000 individuals, in Iran of 80:100,000 inhabitants [Davatchi *et al*., 2008] and in Portugal, of 2.4 per 100,000 inhabitants [Grupo Nacional para o Estudo da Doença de Behçet *et al*., 1997], one of the highest prevalence among occidental European countries.

The age at onset of BD is most commonly in the third decade, although the age at the time of final diagnosis usually is in the fourth decade [Bang *et al*., 1997]. Both genders are affected, although both male and female predominance have been described in different geographic regions. In eastern Mediterranean populations, the disease is more common in

men, who also experience more severe disease [Tursen *et al*., 2003]. In Asian populations, however, the sex ratio is reversed [Bang *et al*., 1997]. In Iran, the male to female ratio has been calculated to be 1.22:1.00 [Davatchi *et al*., 2010].

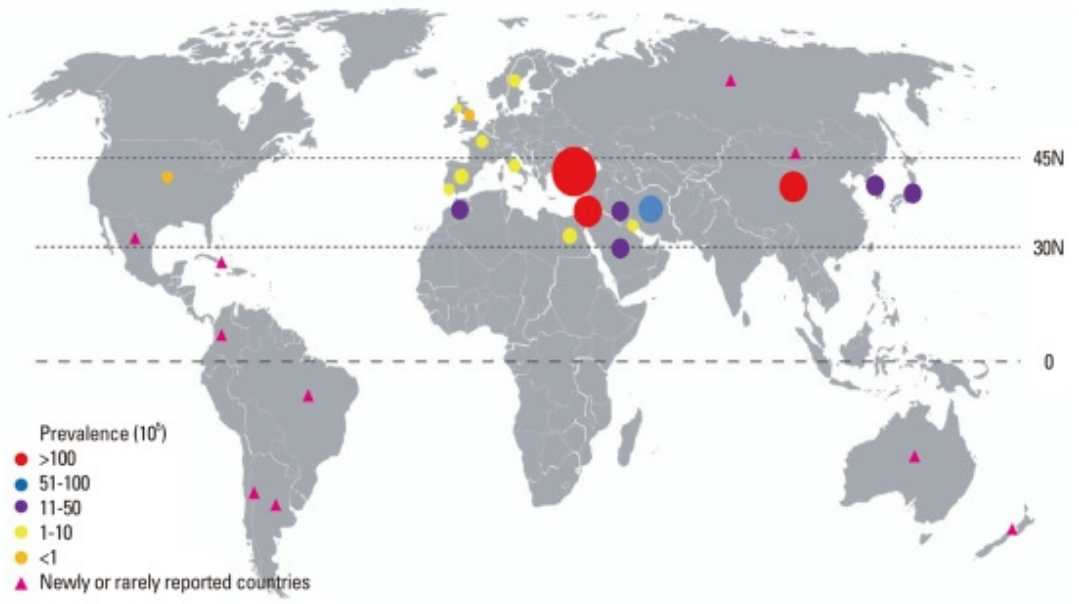


Figure 1. Global distribution of Behçet's disease. Dot size reflects prevalence. Figure taken from Cho *et al*., 2012.

About 2-3% of all affected individuals have childhood-onset BD (generally defined as definitive BD before the age of 16 years) [Saylan *et al*., 1999] and most affected children develop clinical symptoms between 7 and 13 years of age [Koné-Paut *et al*., 1998; Gül *et al*., 2000, Koné-Paut *et al*., 2002].

Most BD cases are sporadic but a higher incidence of familial aggregation has been observed in paediatric BD patients and in up to 12% of non-caucasian patients [Koné-Paut *et al*., 1999]. A sibling risk ratio of 11.4-52.5 has in fact been reported in a Turkish study, suggesting a strong genetic background [Gül *et al*., 2000]. Furthermore, an increase prevalence of isolated manifestations of disease has also been observed among first degree relatives of the patients [Chamberlain, 1978; Bird Stewart, 1986].

Although it is clear that there is a significant genetic component in the susceptibility to Behçet's disease, environmental factors also play a role. The study of migrant populations has yielded interesting epidemiological findings: Turkish individuals who have migrated to Germany have a significantly lower risk of disease than individuals of Turkish origin living

in Turkey, although their risk remains higher than that of the native German population [Zouboulis *et al*., 1997]. Similarly, the disease is virtually unknown in Japanese immigrants to Hawaii, mainland USA or South America [Sakane *et al*., 1999] despite a high prevalence in Japan.

1.1.3 Clinical features

BD is classified among vasculitides, but its clinical picture is very distinctive from the other vasculitides. It gives rise to an ample spectrum of clinical features characterized by unpredictable exacerbations and remissions. Its major symptoms are the recurrent orogenital ulceration, as well as the eye and skin lesions [reviewed in Marshall, 2004]

Recurrent oral aphthosis

Recurrent oral aphthosis (ROA) is the sine qua non of Behçet's disease. Oral aphthae are observed in nearly every patient and are usually the earliest sign of disease [reviewed in Marshall, 2004]. They may precede the onset of systemic symptoms by many years [reviewed in Marshall, 2004] and isolated ROA is often reported in family members of BD patients [Krause *et al*., 1999].

Genital aphthosis

Genital aphthosis (GA) occurs in 72-94% of BD patients and is morphologically similar to oral aphthous [reviewed in Marshall, 2004]. GA is rarely the initial manifestation of BD and is less common than oral aphthosis. Even though arising in both children and adults, they are more common in the latter [Krause *et al*., 1999; reviewed in Al-Otaibi *et al*., 2005]

Ocular disease

Most studies report ocular disease in 30-70% of patients with BD and it seems to be more common and severe in men than in women [Kural-Seyahi *et al*., 2003; Tursen *et al*., 2003]. Ocular disease is usually bilateral and characteristically occurs within 2–3 years of disease onset [Kural-Seyahi *et al*., 2003]. Anterior uveitis (AU) is a characteristic sign of BD, and together with posterior uveitis (PU) and retinal vasculitis (RV), may lead to partial

visual loss in up to 25% of the patients, or even in some cases to complete blindness, although the prognosis is improving with the use of modern immunossupressors [Muhaya *et al*., 2000].

Skin lesions

Skin manifestations occur in about 80% of BD patients. The most frequent lesions are pseudofolliculitis (PF) and erythema nodosum (EN). These lesions usually arise simultaneously, most frequently on the skin of the back, face, and chest [reviewed in Marshall, 2004].

Vascular manifestations

Behçet's disease is a chronic relapsing systemic vasculitis, affecting arteries and veins of all sizes. The vascular manifestations (VMs) may consist of superficial phlebitis, deep vein thrombosis, large vein thrombosis, arterial thrombosis, and aneurysm. Venous involvement usually arises within five years of the initial presentation of BD. Small-vessel vasculitis accounts for much of the pathological process of disease, and clinically manifest large-vessel involvement occurs in between 7 and 49% of patients [reviewed in Al-Otaibi *et al*., 2005].

Articular disease

Joint manifestations affect almost two thirds of BD patients, the most frequent manifestation being non-erosive and non-deforming oligoarthralgia that typically involves the knees, ankles and wrists. The arthritis of BD is usually transient, being rare the long-term arthritic disease [reviewed in Marshall, 2004].

Neurological involvement

Involvement of the central nervous system (CNS) that occurs in 5–10% of patients is more common than involvement of peripheral nerves, although, rarely, they may also be involved. The neurological manifestations usually occur within 5 years of disease onset and are most common in men. Neurological disease carries a high morbidity and the mortality is

usually estimated at 5–10% [Akman-Demir *et al*., 1999, Koné-Paut *et al* 2002; reviewed in Marshall, 2004].

Gastrointestinal lesions

The frequency of gastrointestinal tract involvement varies considerably in different populations. In Iran it has been estimated to affect 7.4% of BD patients [Davatchi *et al*., 2010b]. The most common gastrointestinal feature of BD is the mucosal ulcers, but the spectrum of clinical symptoms is wide and includes anorexia, vomiting, dyspepsia, diarrhoea and abdominal pain [reviewed in Marshall, 2004].

Pathergy

The pathergy phenomenon (PP) is a non-specific hyperactivity of the skin, in response to minor trauma that occurs in Behçet's disease. The formal pathergy test involves the intradermal insertion of a 20-gauge needle under sterile conditions and without injecting saline. It is considered positive if an erythematous sterile papule develops within 48 hours. The variable association of pathergy test positivity with BD and its occurrence in healthy individuals prevent its use as a screening test [reviewed in Marshall, 2004].

1.1.4 Diagnostic criteria

There is no specific test for Behçet's disease, and the diagnosis is based upon clinical criteria. A number of different diagnosis/classification criteria have been proposed, each using different sets of clinical features [reviewed in Marshall, 2004]. In 1990, the International Study Group Criteria for Behçet's Disease (ISG) (Table 1) [ISG, 1990] were developed by an international group of experts. These criteria have been widely adopted, and although proposed for the definition of patients participating in research studies, they also perform well in the clinical context of patient's diagnosis [reviewed in Al-Otaibi *et al*., 2005]. These ISG criteria require the presence of recurrent OA (at least three times in one year) plus two of the following symptoms: recurrent GA, eye lesions (AU, PU or RV), skin lesions (EN or PF), or a positive PP (Table 1). The criteria have excellent specificity (measured by the percentage of healthy individuals, correctly recognized as not having BD), however they lack sensitivity

(measured by the proportion of BD patients correctly classified by the criteria) [Davatchi, 2012].

Table 1. International study group (ISG) criteria for the diagnosis of Behçet's disease [ISG, 1990].

Symptoms	Description
1. Recurrent oral ulceration	Minor aphthous, major aphthous or herpetiform ulceration observed by physician or reported by patient, which have recurred at least three times in a 12-month period.
And two of the following:	
2. Recurrent genital ulceration	Recurrent aphthous ulceration or scarring, observed by physician or reported by patient.
3. Eye lesions	Anterior uveitis or posterior uveitis, cells in vitreous on slit lamp examination or retinal vasculitis observed by ophthalmologist.
4. Skin lesions	Erythema nodosum lesions observed by physician or patient, pseudofolliculitis or papulopustular lesions or acneiform nodules consistent with disease observed by the physician.
5. Positive pathergy test	Read by physician at 24–48 hours.

In 2006 the International Criteria for Behçet's Disease (ICBD) (Table 2) were established [International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD), 2006]. ICBD uses six items: OA, GA, skin lesions (PF, EN), eye lesions (AU, PU, RV), VMs and PP. Genital aphthosis and eye lesions have more diagnostic value than the others symptoms and each receive 2 points. The other 4 items (OA, skin lesions, VMs and PP) correspond to one point each. A patient that has 3 or more points is diagnosed as having BD [Davatchi, 2012].

Table 2. Revised International Criteria for Behçet's disease (ICBD) [ITR-ICBD, 2006].

Symptoms	Points
1. Oral aphthosis	1
2. Genital aphthosis	2
3. Eye lesions	2
4. Skin lesions	1
5. Positive pathergy test	1
6. Vascular lesions	1

Although the diagnosis of Behçet's disease may be straightforward once the possibility has been recognized, incomplete disease or unusual presentations often represent a diagnostic challenge. A detailed clinical history is essential to exclude other conditions and reveal subtle features of this complex disease [reviewed in Marshall, 2004].

1.1.5 Histopathology

BD lesions are characterized by an early neutrophil infiltration, monocytic cellular infiltration and endothelial cell swelling with or without fibrin deposition in the vessel wall and necrosis of the surrounding tissue [reviewed in Mendoza-Pinto *et al*. 2010] Neutrophil infiltration occurs with variable degrees of vascular infiltration particularly in early lesions including those of the skin pathergy reaction, mucocutaneous aphthae, nodular cutaneous lesions, and ocular lesions. Interaction of cellular adhesion molecules together with endothelial proliferation may play an important role in the formation of skin pathergy reaction lesions in patients with BD [Köse, 2012]. Perivascular infiltration of memory T cells and polymorphonuclear leucocytes within vasculitic lesions in BD patients who have arterial and CNS involvement have also been reported [Hirohata, 2008].

1.1.6 Etiopathogenesis

The cause of BD remains uncertain, but the most likely hypothesis is that an autoimmune and/or inflammatory reaction is triggered by an infectious or environmental agent in a genetically predisposed individual, resulting in an inflammation of the affected organs mediated by the immune system [Kapsimali *et al*. 2010]. Therefore, complex interactions between environmental, genetic and immunological factors are believed to contribute to the pathogenesis of BD. Most recent studies have focused on new genetic markers and in the involvement of T cells.

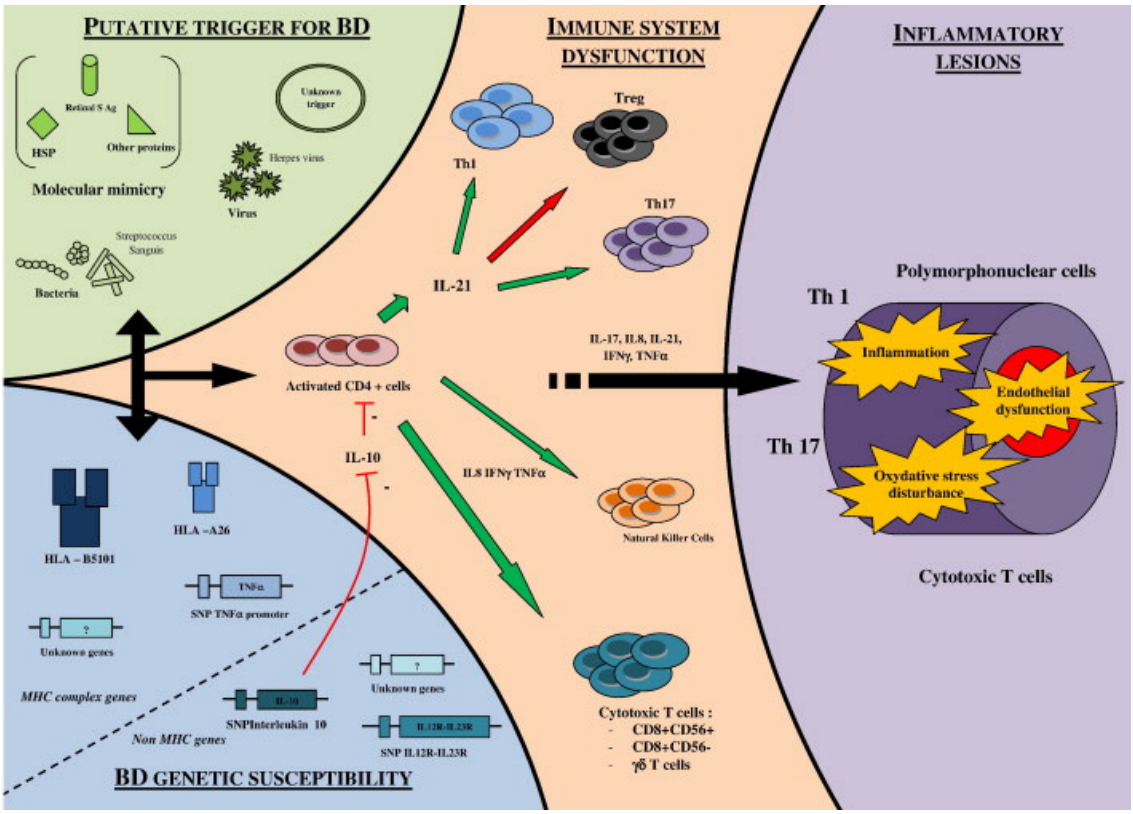


Figure 2. Current model for Behçet's disease pathogenesis. Figure taken from Pineton de Chambrun *et al.*, 2012.

1.1.7 Suggested environmental trigger

Several studies have suggested a variety of infectious agents which may trigger an immune response, thus inducing immunological cascade that gives rise to an exacerbation of the disease.

The most plausible environmental trigger is an infectious agent, and evidence of ongoing or previous infection with a variety of viral agents has been investigated. These include herpes simplex virus [Eglin *et al.*, 1982; Studd *et al.*, 1991; Nomura *et al.*, 1998] since serum antibodies against HSV-1 and circulating immune complexes are found in patients with BD, the hepatitis viruses [Aksu *et al.*, 1999; Ilter *et al.*, 2000] and the parvovirus B19 [Kiraz *et al.*, 2001]. Potential bacterial triggers include mycobacteria [Lehner *et al.*, 1991] and Helicobacter pylori [Avci *et al.*, 1999]. However the most commonly investigated microorganism in the pathogenesis of BD is Streptococcus. The relationship between streptococcal infections and BD is suggested by clinical observations such as a higher incidence of nasopharyngeal infections and dental caries, aggravation of BD by dental

treatment and the beneficial effect of antibacterial treatments on mucocutaneous and synovial disease symptoms [Mumcu *et al*., 2007].

However, there is still no consistent evidence supporting the role of a single microorganism in the aetiology of BD [Kapsimali *et al*., 2010]. It is possible that it is not a specific microorganism but its presence and persistence that might determine its role in BD pathogenesis but remains unclear how can the microbial antigens activate the immune system and lead to the vasculitic mechanism observed in BD.

1.1.8 Immunological mechanisms

Alternatively, Behçet's disease may be primarily autoimmune in origin. This does not exclude an external agent, which could operate through molecular mimicry [Sakane *et al* 1997] or some other mechanism, but implies that the disease is perpetuated by an abnormal immune response to an auto-antigen in the absence of ongoing infection [Benoist *et al*., 2001, reviewed in Marshall, 2004].

To date, neither *in vitro* nor *in vivo* models of the immunopathogenesis of BD remain and a multi-systemic animal model of BD is also non-existent. Furthermore, BD doesn't have classical autoimmune diseases features such as female preponderance, association with other autoimmune diseases, specific associated auto-antibodies or antigen-specific T cells [Yazici, 1997; Direskeneli, 2006]. Also, most autoimmune diseases are shown to have an major histocompatibility complex (MHC) class II association, however BD is associated with a MHC class I allele, *HLA-B*51*. This association is similar to spondylarthropathies, that are associated with other MHC class I allele, *HLA-B*27* [Direskeneli, 2006]. In common with autoimmune diseases, BD exhibit favourable responses to immunomodulatory or immunosupresive treatment. However tumor necrosis factor alpha (TNF-α) antagonists are also an exception in this concept, as they act mainly as anti-inflammatory agents and are accepted to be contraindicated in autoimmune diseases such as systemic lupus erythematosus (SLE) and multiple sclerosis but are effective in BD treatment [Direskeneli, 2006] (please refer to chapter 1.1.10.1, page 27, for a short description on the therapeutics used in BD treatment).

BD also have characteristics that fits the spectrum of auto-inflammatory disorders such as the episodes of inflammatory attacks of innate nature with neutrophil-dominated infiltrates and with over-expression of pro-inflammatory cytokines, the recurrent non-scaring mucocutaneous lesions and non-deforming arthritis [Gül, 2005]. Therefore, the

activation of the adaptative immune system (with reaction against self-antigens) can also be a secondary response to a profound and uncontrolled innate-related inflammatory reaction [Gül, 2005; Marshall, 2004] and not be in the origin of the vasculitis symptoms. In either way, the innate and adaptative systems are integrated in BD in a close that still needs to be clarified.

1.1.8.1 Autoantibodies

Autoantibodies are frequently reported in autoimmune diseases, with a demonstrated pathogenic role for some of them. Antibodies against cell surface such as the endothelial cell antibodies (AECE), that are a frequent but non-specific finding in BD patients, have been proposed to play a role in BD pathology. Antibodies against α-tropomyosin [Matsui *et al*., 1999] and kinectin were also shown. However, although there is evidence of an inflammatory response against some autoantigens in BD, there are no conclusive and convincing results implicating autoantibodies in BD pathology [Pineton de Chambrun *et al*., 2012].

1.1.8.2 Heat shock proteins

Heat-shock proteins (HSPs) are a group of intracellular proteins which scavenge for other intracellular proteins when eukaryotic and prokaryotic cells are exposed to denaturating stress conditions (infection, trauma, heat, hypoxia, and cold) [Lamb *et al*., 1990; Direskeneli *et al*., 2003]. A statistically significant increase in the expression of HSP60/65 was noted in patients with active BD and mucocutaneous ulcers compared to controls [Ergun *et al* 2001]. Significant sequence homology exists between mammalian and microbial HSPs. It is hypothesize that cross-reaction (through molecular mimicry) between peptides from bacterial or viral HSP, homologous human HSP, and mucosal antigens may drive the selection of auto-reactive T-cells, thus linking infection with autoimmunity [Lehner, 1997; Direskeneli *et al*., 2003].

1.1.8.3 Neutrophilic hyperfunction

The involvement of neutrophils in the pathogenesis of BD has been extensively studied. These cells are the main elements of the innate immunity and their infiltration plays

a pivotal role in the inflammatory vasculitis observed in BD. The activation of neutrophils has been observed in BD in *vivo* [Eksioglu-Demiralp *et al*., 2001]. Neutrophils hyperactivation is responsible for the endothelial adhesion, chemotaxis, phagocytosis and also for the generation of reactive oxygen species (ROS) which are increased in BD, and that play also an important role BD pathogenesis. On the other hand, cytokines and chemokines secreted from antigen presenting cells (APCs) and T cells are suggested to cause neutrophil hyperactivation. Activated neutrophils secrete some cytokines which prime themselves and also cause T helper cells type-1 (Th1) stimulation, therefore, the crosstalk between APCs, Th1 lymphocytes and neutrophils has been proposed to be a critical parameter of the immune-mediated mechanism [Zouboulis *et al*., 2000; Yazici, 2004; Katsantonis *et al*., 2000; reviewed in Pay *et al*., 2007]. It is still unclear whether the neutrophil hyperactivation observed in BD patients reflects genetic influences or persistent activation by external agents [Takeno *et al*., 1995].

1.1.8.4 Abnormal T cell response

There is a strong evidence for generalized aberrant T-cell responses in Behçet's disease. When compared with controls, patients with Behçet's disease have an increased number of gamma delta (γδ) T-cells in circulation and at sites of inflammation [Freysdottir *et al*., 1999; Hamzaoui *et al*., 1994]. These γδ T-cells are found in epithelial surfaces where they play a major role in the first line of host defence. They participate in early responses against microorganisms by responding to non-peptide antigens in a non-MHC restricted fashion, exhibiting early activation markers and producing inflammatory cytokines [Freysdottir *et al*., 1999]. Although their target antigen is unclear, experimental models have shown that γδ T-cells are strong inducers of Th1 and Th17 cells. Whether the expansion of γδ T-cells in BD is a primary pathogenic event or an epiphenomenon remains to be elucidated [Kapsimali *et al*., 2010].

An imbalance between T helper cells type-1 and type-2 has been reported in the pathogenesis of several autoimmune diseases. Th1 and Th2 cells have different cytokine secretion patterns and functions. Th1 cells secrete interferon gamma (IFNγ), interleukin (IL) type 2 (IL2), and tumor necrosis factor beta (TNFβ), activate macrophages and promote cell-mediated immunity, whereas Th2 cells produce anti-inflammatory cytokines like IL4, IL5, IL10 and IL13 that inhibit macrophage functions and are involved in antibody-mediated immunity [reviewed in Frassanito *et al*., 1999]. Previous studies have suggested that BD is

predominated by a Th1 immune response, since an increase in Th1 cytokine production has been observed in active BD patients [reviewed in Pineton de Chambrun *et al*., 2012], and circulating T-cells are predominantly of the Th1 phenotype [Frassanito *et al*., 1999]. Levels of IL12, which drives the Th1 response in naïve T cells, were also observed increased in BD patients and correlated with the levels of Th1 lymphocytes [Frassanito *et al*., 1999].

T helper cells type-17, characterized by their production of IL17, have been recently isolated and implicated in many autoimmune/inflammatory diseases. IL17 promotes neutrophil influx and regulates neutrophil-mediated inflammatory responses [Direskeneli *et al*., 2011]. IL23 induces T cell activation to produce IL17 and therefore is one of the main Th17 pathway activator [reviewed in Pineton de Chambrun *et al*., 2012]. Interestingly, high levels of IL17 and IL23 have been reported in peripheral blood mononuclear cells (PBMCs) from active BD patients [Kim *et al*., 2010] and also a high Th17/Th1 ratio was found increase in cerebrospinal fluid (CSF) fluid of nero-Behçet's patients when compared to controls [Hamzaoui *et al*., 2011]. It has been suggested therefore that Th17 cells should be interpreted in the context of their ratio to Th1 cells [Kim *et al*., 2010].

Regulatory T cells (Tregs) constitute a largely heterogeneous population of cells that can be defined based on their potential to suppress other immune cells and thereby limit or suppress immune responses. Tregs have been, therefore, widely studied in different autoimmune disorders and changes of their proportion in peripheral blood (PB) or defect of suppressive function *in vitro* have been reported in these group of diseases [Amelsfort *et al*., 2004; Crispin *et al*., 2003]. Concerning BD different observations have been made in respect to these cells: some studies have reported an increase of Tregs in PB and CSF of BD patients in the active phase as compared with BD in remission and healthy controls [Hamzaoui *et al*., 2006; Hamzaoui *et al*., 2007]; a different study has reported a lower level of these cells before ocular attacks in BD patients, than afterward speculating that this could reflect a trafficking of these cells to the eyes [Nanke *et al*., 2008]; but a decrease of total Tregs and active Tregs in PB from BD patients compared to controls was also observed [Kim *et al*., 2012]. These conflicting results may be due to differences in the detection markers, parameters tested (e.g., measuring of total Tregs versus measuring of active and resting Tregs) and applied methodology [Kim *et al*., 2012]. More recently it was reported a marked increase in Th17 and Th1 cells and a decrease in the frequency of Tregs in peripheral blood that were induced by IL21 production and that correlate with BD activity. Conversely, IL21 blockade restored the Th17 and Treg homeostasis in patients with BD [Geri *et al*., 2011]. In the same work authors also demonstrated the presence of IL21 and IL17A producing T cells within the CSF, brain

parenchyma inflammatory infiltrates, and intracerebral blood vessels from patients with active BD and central nervous system involvement. They suggested a two-step model in which Th17 cells operate first leading to the recruitment in the CNS of second wave of T cells including Th1 and inflammatory leucocytes [Geri *et al*., 2011], establishing a link between innate and adaptative immune responses in BD patients [Geri *et al*., 2011; Direskeneli *et al*., 2011].

1.1.8.5 Cytokines

Overproduction of activated monocyte-derived pro-inflammatory cytokines (e.g., IL1, IL6, IL8 and TNF-α) has been implicated in the BD pathogenic process, and their increased levels may represent a disease activity marker [Zouboulis *et al*., 2000; Katsantonis *et al*., 2000; Zouboulis *et al*., 2003]. An elevation of IL10 can also occur in BD and IL12, a potent immunoregulatory cytokine was found to be elevated in peripheral blood of BD patients and correlated with disease activity [Turan *et al*., 1997; Frassanito *et al*., 1999]. In addition, it seems that IL18 and IFNγ contribute to the mucocutaneous inflammatory reaction [Ben Ahmed *et al*., 2004].

1.1.9 Genetic factors

A genetic basis for Behçet's disease is suggested by the disease peculiar ethnic distribution, the observed association with certain human leucocyte antigen (*HLA*) antigens and familial aggregation of some affected individuals. However, the inheritance of Behçet's disease does not follow Mendelian patterns and most cases are sporadic although the familial occurrence of the disease has been reported in up to 12% of non-caucasian patients [Koné-Paut *et al*., 1999]

1.1.9.1 HLA region

The *HLA-B*51*, shows the strongest evidence of association with BD described to date [Ohno *et al*., 1982]. The MHC region, where *HLA-B*51* resides, is a highly polymorphic region located on chromosome 6p21.33, that encompasses ~4000kb and has important biological functions (i.e., immune responses, development) and medical impact. This region is divided into three sub-regions, namely *HLA* class II (1Mb), class III (1Mb) and class I (2Mb)

(Figure 3). These genes encode cell-surface molecules that present antigenic peptides to T cells, thereby initiating the immune response against invading pathogens and other foreign antigens. However, the classical *HLA* loci represent a minority of the genes found in the MHC region, since at least another 120 genes are present [Beck *et al*., 2000].

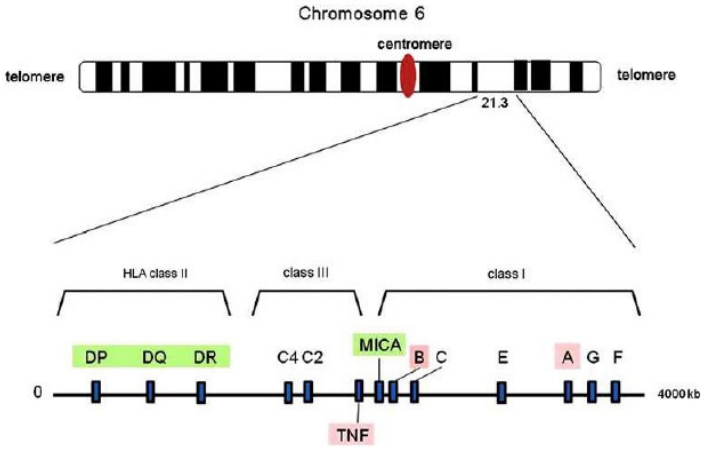


Figure 3. Chromosome map of 6p21.3, which contains the MHC region. The MCH regions span over 4000Kb and consists of the human leukocyte antigen (*HLA*) class II, III and I regions, extending from the centromere to telomere. The following genes are shown: HLA class II genes - *DP*, *DQ*, *DR*, HLA class III genes – Complement component 4 (*C4*) and 2 (*C2*), Tumor necrosis factor (*TNF*), MHC class I chain-related gene (*MICA*), *HLA* class I genes - *B*, *C*, *E*, *A*, *G* and *F*. Figure taken from Song *et al*., 2012.

Ohno *et al* was the first describing the association of BD with *HLA-B*5* in a Japanese population in 1982 [Ohno *et al*. 1982] (for a revision on association studies, please refer chapter 1.2.2, page 30). This association, that was later more specifically linked to its most common split antigen *HLA-B*51*, has been repeatedly demonstrated in different populations [Baricordi *et al*., 1986; Al-Rawi *et al* 1986; Arber *et al*., 1991]. *HLA-B*51* has at least 34 allelic variants but the association has been refined to the most common molecular subtypes, *HLA-B*5101* and *HLA-B*5108* [Kötter *et al*., 2001]. BD cases harbouring *HLA-B*51/B*5* (i.e., *HLA-B*5* or *HLA-B*51*, depending on the allele genotyped) range between 50% and 72% and the pooled risk (estimated based on the odds ratio) of *HLA-B*51/B*5* carriers developing BD has been estimated to be increased by a factor of 5.78 (95% CI of 5.00–6.67). Although the frequency of *HLA-B*51/B*5*-positive BD cases varies across geographic locations, the relative risk associated with this allele appeared to be similar for different ethnic groups [de Menthon *et al*., 2009].

However *HLA-B*51* alone is neither necessary nor sufficient to cause BD. Family studies have estimated that *HLA-B*51* explains only 19% of BD genetic risk [Gül *et al*., 2001], but in a recent meta-analysis of 4,800 BD patients and 16,289 controls from 78 independent studies, its contribution has been estimated to be 32-52% [de Menthon *et al*., 2009]. This data together with the observation that *HLA-B*51* is present at a high frequency in some populations in which the disease is virtually unknown [Verity *et al*., 1999a], indicates that other genetic factors must be involved.

The biological mechanisms through which specific *HLA-B* alleles confer risk for BD remain unknown. It is still not clear whether these are primary associations or whether *HLA-B*51* is associated with BD only because of linkage disequilibrium (LD) with other, causative polymorphism. Candidates include MHC class I chain-related gene A (*MICA*), tumor necrosis factor (*TNF*) and *HSP* genes which are in LD with *HLA-B*51*. In fact, several polymorphisms in these genes have been identified to be associated with BD, although their role in BD pathogenesis is not definitive [Direskeneli, 2001; Ahmad *et al*., 2003; Mizuki *et al*., 2007]. Independent associations from *HLA-B*51* at the *MHC* locus include: *HLA-A*26* that was found associated in Greek and Japanese patients [Kera *et al*., 1999; Meguro *et al*., 2010], *HLA-B*15* that was found associated in Moroccan BD patients but was not replicated in other datasets [Choukri *et al*., 2001], *HLA-B*5701* found associated in patients from Great Britain and not replicated as well [Ahmad *et al*., 2003] and *TNF-a-1031C* found associated in different populations [Ahmad *et al*., 2003; Park *et al*., 2006]. The *MICA* was considered a major candidate for being responsible for BD genetic susceptibility due to its localization 46Kb centromeric to *HLA-B* and to its immunological function as a stress-inducible antigen involved in recognition and interaction with natural killers (NK) and γδ, alpha beta CD8+ T cells (αβ T-cells) [Russell *et al*., 2001]. However, it is now accepted that the association of *MICA* is attributed to its strong LD with *HLA-B*51* [Piga *et al*., 2011].

Cross-reactivity to self-antigens (e.g., retinal S), as well as interactions between killer cell immunoglobulin-like receptors (KIRs) expressed on NK and γδ T-cells and *HLA-B*51*, have been proposed as potential pathogenetic mechanisms [Direskeneli, 2001]. However, the most compelling theory is that *HLA-B*51* presents autoantigens to T-suppressor lymphocytes and thereby activates the immune system. It remains however theoretically possible that a particular isoform of *MICA* might act in combination with *HLA-B*51* to increase the risk or severity of the disease [Russell *et al*., 2001]. The *HLA-B*51* molecule itself may also be responsible, at least in part, for neutrophil hyperfunction in BD, since HLA-B*51-transgenic mice show enhanced neutrophil function as seen in BD patients, although these mice did not

develop the symptoms of BD [Takeno *et al*., 1995]. Since the primary role of HLA class I antigens such as *HLA-B*51* is to present endogenous peptides to CD8+ T cells, the lack of the disease phenotype in this mouse model can be explained by the absence of an triggering microbial or injury-related peptide that would activate the disease-relevant CD8+ T cells [Kaya, 2012].

The variable nature of clinical BD manifestations raised also the question of whether *HLA-B*51/B*5* could have a modulatory effect on disease expression. Past studies suggested that *HLA-B*51/B*5*-positive and negative BD patients differed in that the former developed more frequently CNS [Sakane *et al*., 1999] or eye manifestations [Verity *et al*., 1999b] and the latter more commonly thrombophlebitis [Azizleri *et al*., 1994]. In addition, it was suggested that patients harbouring the *HLA-B*51/B*5* allele have more unfavourable BD courses, characterized by poorer outcomes of ocular [Sakane *et al*., 1999; Kim *et al*., 1989; Soylu *et al*., 1992] or neurological involvement [Sakane *et al*., 1999]. However, these observations were not consistently reported among studies and the discrepancies may have been exacerbated by studies with small sample sizes. A recent meta-analysis performed by Maldini *et al*., (2012) indicate that the *HLA-B*51/B*5* allele increases the risk for genital ulcers, ocular involvement and skin manifestations, decreasing the risk of gastrointestinal involvement, but with only a modest effect. This implies that the clinical pictures of *HLA-B*51/B*5*-positive and -negative BD patients are virtually indistinguishable, and that genotyping of this allele cannot accurately predict the occurrence of organ specific or systemic manifestations [Maldini *et al*., 2012]. Another finding of Maldini *et al.,* (2012) was that *HLA-B*51/B*5* carriage is more common in male BD patients [Maldini *et al*., 2012], consistent with previous observations [Choukri *et al*., 2001; Kilmartin *et al*., 1997; Lehner *et al*., 1979]. Their results also reinforced previous results suggesting that the strength of the *HLA-B*51/B*5* BD relationship was positively correlated with the proportion of male cases included in individual studies [de Menthon *et al*., 2009].

In a recent study Hughes *et al*., (2013) performed an extended genotyping in the *HLA* region (8572 single nucleotide polymorphism (SNPs) genotyped using a custom platform - Immunochip), in a Turkish dataset of 503 BD cases 504 controls and in an Italian dataset of 144 BD cases and 1270 controls. The genotyping was followed by imputation (for an introduction to these method please refer to chapter 1.2.2.4, page 43) and association testing of both genotyped and imputed SNPs in the two datasets (using data from the 1000 Genomes Project). The genotyped SNPs were used to infer classical *HLA* alleles in the *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1* loci. Their results suggested that *HLA-B*51*

association in BD is explained by a variant located in the promoter region of *HLA-B*, between *HLA-B* and *MICA* genes. This variant (rs116799036, $P_{meta\text{-}analysis}$=9.42E-50, odds ratio (OR) = 3.88, 95% confidence interval (CI) not shown) remained significantly associated with BD independently of *HLA-B*5101* and notably the association of *HLA-B*5101* with BD was completely abrogated after adjusting the analysis on this variant ($P_{adjusted}$ =1.60E-01) [Hughes *et al*., 2013]. This study also identified an independent association in the psoriasis susceptibility 1 candidate 1 (*PSORS1C1)* loci (rs12525170, $P_{meta\text{-}analysis}$=3.01E-26, OR=3.01, 95%CI not shown), in the *HLA-Cw*1602* allele (OR=5.38, P-value = 6.07E-18) and in one variant located in the *HLA-A* region, upstream of *HLA-F-AS1* (rs114854070, $P_{meta\text{-}analysis}$=7.84E-14, OR=1.95, 95%CI not shown). The previous reported association with *HLA-A*26* was not tested by the authors due to the low frequency of this allele in both populations [Hughes *et al*., 2013]. This study suggests that the risk previously ascribed to *HLA-B*51* is likely not causal in BD and highlight the importance of re-evaluate classical *HLA* association in other diseases using a more advance panel of markers and statistical approaches.

1.1.9.2 Linkage study

Please refer to chapter 1.2.1 for an introduction to linkage studies. Karasneh and his colleagues performed the only published whole genome linkage analysis in BD using a set of 395 microsatellites (MSs) in 28 Turkish multicase families. The study identified 16 potential loci for BD at chromosomes 1p36, 4p15, 5q12, 5q23, 6p22–24, 6q16, 6q25-26, 7p21, 10q24, 12p12-13, 12q13, 16q12, 16q21–23, 17p13, 20q12-13, and Xq26–28 with the strongest evidence seen for 12p12-13 and 6p22–24 [Karasneh *et al*., 2005a].

1.1.9.3 Genome-wide association studies

Please refer to chapter 1.2.2.2 for an introduction to genome-wide association studies (GWASs). Six GWASs for Behçet's disease have been published until May 2013. The main findings of each study are summarized in Table 3.

The first GWAS for Behçet's disease was performed in a relatively small sample of 152 BD patients and 172 controls of Turkish origin using Affymetrix 500K arrays [Fei *et al*., 2009]. Fei and colleagues identified genetic associations between BD and SNPs in chromosome 9 open reading frame 174 (*KIAA1529*) (rs2061634, *P*=4.20E-04, OR[95%CI] = 2.04[1.45-2.88]), carboxypeptidase, vitellogenic-like *(CPVL*) (rs317711, *P*=1.00E-04,

OR[95%CI] = 2.26[1.47-3.45]), hypothetical protein LOC100129342 (rs11206377, *P*=3.00E-04, OR[95%CI] = 1.84[1.32-2.58]), ubiquitin-associated and SH3 domain containing, B *(UBASH3B*) (rs4936742, *P*=1.50E-03, OR[95%CI] = 1.71[1.23-2.38]), and ubiquitin-associated domain containing, 2 (*UBAC2*) (rs9513584, *P*=5.80E-03, OR[95%CI] = 1.61[1.15-2.26]) genes, in a Turkish dataset but independent replication was not attempted in this study. Interestingly, none of these five susceptibility loci had been associated with BD before. The functions of *KIAA1529* and *LOC100129342* are not known, but the genes *UBASH3B* and *UBAC2* both contain a UBA, suggesting that both gene products are involved in the ubiquitination pathway. The *CPVL* gene encodes for a carboxypeptidase that cleaves a single amino acid from the carboxy terminus of proteins or peptides [Fei *et al*., 2009].

The second GWAS was performed using a panel of 23,465 MSs in a Japanese population of 300 cases and 300 controls. Meguro and his team identified 6 positively associated microsatellites (D3S0186i, D6S0014i, D6S0032i, 536G12A, D12S0645i, and D22S0104i) [Meguro *et al*., 2010]. The markers 536G12A and D12S0645i are located in two BD linkage regions (6q25.1 and 12p12.1) identified by Karasneh and his colleagues in 2005 [Karasneh *et al*., 2005a]. However more studies are needed to identify the importance of these microsatellites in the pathogenesis of BD. They also performed a HLA class I analysis and identified the *HLA-B*5101* as the strongest susceptibility allele (OR=5.50, 95%CI not shown) and an independent association at *HLA-A*2601* that was stronger in *HLA-B*51* negative patients ($P_{corrected}$=0.016 and OR [95%CI] = 1.92 [1.31-2.83] when considering all patients and $P_{corrected}$=5.11E-04 and OR [95%CI] = 2.61 [1.66-4.10] when considering only *HLA-B*51* negative patients) [Meguro *et al*., 2010].

Remmers *et al*., (2010) reported a GWAS with 311,459 SNPs in 1215 BD patients and 1278 controls from Turkey. They confirmed the *HLA-B*51* association and identified a second, independent association within the *HLA-A* region (rs9260997, *P*=5.49E-09, after conditioning the analysis on *HLA-B*51*). They also identified near genome-wide significance for one non-MHC SNP, rs936551 (*P*=5.29E-08) located in complexin-1 gene (*CPLX1),* and strong suggestive association for rs3024490 (*P*=2.22E-07), located in *IL10*. They also fine-mapped the region of the interleukin 23 receptor gene *(IL23R)* and interleukin 12 receptor, beta 2 subunit gene (*IL12RB2*), since this region was identified as a susceptibility locus in an independent GWAS performed in the Japanese population and published back-to-back in Nature Genetics [Mizuki *et al*., 2010]. The results of the meta-analysis including a total of 2430 cases and 2660 controls established associations with the *IL10* gene variant (rs1518111,

P=3.54E-18, OR [95%CI] = 1.45 [1.34-1.58]) and with a variant located between the *IL23R* and *IL12RB2* genes (rs924080, *P*=6.69E-09, OR [95%CI] = 1.28 [1.18-1.39]) [Remmers *et al*., 2010].

In the Mizuki *et al*., (2010) GWAS, 500,568 SNPs were investigated in 612 Japanese individuals with BD and 740 unaffected controls. The HLA-B region showed the most significant association with BD (rs4959053, *P*=1.80E−26) and the authors also reported an independent association at the *HLA-A* region. Outside the *HLA* complex, they identified 54 SNPs with a *P*-value <1.00E-04. Genome-wide significant associations (*P*<5.0E-08) were detected and replicated in an intergenic region between *IL23R* and *IL12RB2* (rs1495965, *P*=1.90E-11, OR[95%CI] = 1.35[1.24-1.47], and the second strongest region of association was within the intron 3 of *IL10* (rs1900872, *P*=2.10E-14, OR[95%CI] = 1.45[1.32-1.59] and rs1800871, *P*=1.00E-14, OR[95%CI] = 1.45[1.32-1.60]) [Mizuki *et al*., 2010]. The results of these two GWASs provide cross-validation of association with BD of both *IL10* and *IL23R-IL12RB2*. IL23 is a heterodimeric pro-inflammatory cytokine [Parham *et al*., 2002] that has been shown to stimulate Th17 cell proliferation and increase the production of inflammatory cytokines such as IL1, IL6, IL17 and TNFα [Langrish *et al*., 2005; Aggarwal *et al*., 2003; Liang *et al*., 2006]. *IL12RB2* encodes an IL12 receptor chain. IL12 plays an important role in Th1 cell differentiation [Manetti *et al*., 1993], T-cell and NK-cell cytotoxicity, and IFNγ production by T cells and NK cells [Iwakura *et al*., 2006]. IL12RB2 has been reported to be essential for high-affinity IL12 binding and IL12 dependent signaling and to be up-regulated by IFNγ in Th1 cells [Chang *et al*., 1999]. IL10 is a potent suppressor of inflammatory cytokines such as IL1, IL6, IL12, TNFα and IFNγ and inhibits the costimulatory activity of macrophages for T-cell and NK-cell activation [Mizuki *et al*. 2010].

In 2012 was performed a GWAS on the Chinese population using a discovery dataset of 149 patients and 951 controls and a replication dataset consisting of 554 BD patients and 1159 Chinese controls. Hou and colleagues identified 31 associated SNPs using a *P*-value<1.0E-04. Replication analysis confirmed the association of 3 linked SNPs in signal transducer and activator of transcription 4 gene (*STAT4)* (rs897200, *P*=6.20E-09, OR[95%CI] = 1.45[1.30-1.60), rs7574070 *P*=3.36E-07, OR[95%CI] = 1.40[1.20-1.60] and rs7572482, *P*=1.30E-08, OR[95%CI] = 1.44[1.30-1.60]). They found an increased expression of *STAT4* in individuals carrying the rs897200 risk genotype AA and also that the IL17 messenger ribonucleic acid (RNA) and protein levels were increased in these individuals, an interesting finding since STAT4 regulates the production of IL17 and interferon-gamma. They suggested a model in which up-regulation of *STAT4* and consequently of pro-inflammatory cytokines such as IL17 constitute a potential pathway leading to BD [Hou *et al*., 2012a].

More recently, a GWAS was performed in the Korean population on 379 BD patients and 800 controls. Lee *et al.*, (2013) identified 55 SNPs with *P*-value<4.61E-06 (corresponding to 5% of false discovery rate), 23 of them on the chromosome 6p21.33. Outside the MHC region, the top five SNPs were located in the intergenic region of the GTPase, IMAP family member 7 (*GIMAP7)* and member 4 *(GIMAP4*) genes (rs11769828, *P*=1.60E-06), at WW domain-containing protein 1 gene (*WWC1*) (rs1965673, *P*=4.91E-07), at YTH domain-containing protein 1 gene (*YTHDC1*) (rs10033058, *P*=2.38E-06) and upstream phospholipase C, beta 1 gene (*PLCB1*) (rs4239774, *P*=2.07E-06) genes. Fine mapping analysis of the *GIMAP* locus and replication analysis in a Japanese population of 363 BD patients and 272 controls confirmed the associations at rs1522596 (*P*=1.84E−04, OR[95%CI] =1.29[1.13-1.48]) in *GIMAP4*; rs10266069 (*P*=6.10E−05, OR[95%CI] = 1.32[1.15-1.51]) and rs10256482 (*P*=1.62E−04, OR[95%CI] =1.30[1.13-1.48]) in *GIMAP2*. Their functional studies revealed that the minor allele of rs1608157 results in significantly lower activity of *GIMAP4* promoter than the major allele. Moreover, they also observed that CD4+ T cells from BD patients showed a lower level of *GIMAP4* messenger RNA (mRNA), and *GIMAP4* knockdown was protective against Fas-mediated apoptosis [Lee *et al*., 2013].

Table 3. Summary of top findings from genome-wide association studies performed for Behçet's disease. Allelic *P*-value in the discovery GWAS samples (P_{allele}) are shown for all top SNPs. Overall meta-analyses results ($P_{meta\text{-}analysis}$) are shown when the SNPs were tested for association in an independent population. MEA means Middle Eastern Population, [a] corresponds to SNPs not genotyped in the first phase of the GWAS, but genotyped in a second phase of fine mapping; [b] corresponds to a surrogate marker for rs936551 since rs936551 failed QC in the replication assay; [c] corresponds to a surrogate marker for Japanese BD GWAS *KLRK1* SNP rs2617150; * corresponds to a *P*-value obtained by logistic regression after conditioning on HLA-B*51; [#] to *P*-values results obtained from combined analysis and not meta-analyses; and [£] refers to *P*-values obtained from the combined dataset of uveitis and non-uveitis patients, under a recessive model of inheritance.

Authors	Cases vs Controls	Discovery Population	Number of markers	Associated markers	Nearest gene	*P* allele	Replication in independent population
Fei *et al*., 2009	152 vs 172	Turkish	500,000 SNPs	rs11206377	*LOC100129342*	3.00E-04	Not associated in Chinese [Hou *et al*., 2012b]
				rs317711	*CPVL*	1.00E-04	Not associated in Chinese [Hou *et al*., 2012b]
				rs2061634	*KIAA1529*	4.20E-04	Not associated in Chinese [Hou *et al*., 2012b]
				rs4936742	*UBASH3B*	1.50E-03	Not associated in Chinese [Hou *et al*., 2012b]
				rs9513584	*UBAC2*	5.80E-03	Replicated in Chinese [Hou *et al*., 2012b]
Meguro *et al*., 2010	300 vs 300	Japanese	23,465 MS	D3S0186i	*ROBO1*	2.90E-02	Not associated in Korean [Horie *et al*., 2012]
				D6S0014i	*FLJ45422*	<1.00E-04	Not associated in Korean [Horie *et al*., 2012]
				D6S0032i	*HLA-B*	<1.00E-04	Replicated in Korean [Horie *et al*., 2012]
				536G12A	*PPIL4*	3.10E-02	Not associated in Korean [Horie *et al*., 2012]
				D12S0645i	*SOX5*	2.40E-02	Not associated in Korean [Horie *et al*., 2012]
				D22S0104i	*IGL@*	1.90E-02	Not associated in Korean [Horie *et al*., 2012]
Mizuki *et al*., 2010	611 vs 737	Japanese	500,568 SNPs	rs12119179	*IL23R-IL12RB2*	2.70E-08	-
				rs1495965[a]	*IL23R-IL12RB2*	1.20E-08	Replicated in Turkish [Mizuki *et al*., 2010] and Iranian [Xavier *et al*., 2012]; not associated in Korean [Mizuki *et al*., 2010]
				rs1554286	*IL10*	8.00E-08	Replicated in Iranian [Xavier *et al*., 2012]
				rs1800872[a]	*IL10*	9.50E-09	Replicated in Turkish and Korean [Mizuki *et al*., 2010]
				rs1800871[a]	*IL10*	9.50E-09	Replicated in Turkish and Korean [Mizuki *et al*., 2010]
				rs4959053	*HLA-B*	1.80E-26	-
Remmers *et al*.,	1215 vs 1278	Turkish	311,459 SNPs	rs924080	*IL23R-IL12RB2*	5.35E-06	Replicated in Japanese [Remmers *et al*., 2010] and Iranian [Xavier

2010							*et al*., 2012]; not associated in Turkish, MEA, Greek, UK, Korean [Remmers *et al*., 2010]
				rs1518111[a]	*IL10*	1.88E-08	Replicated in MEA, Greek, UK, Japanese [Remmers *et al*., 2010] and Iranian [Xavier *et al*., 2012]; not associated in Turkish and Korean [Remmers *et al*., 2010]
				rs3024490	*IL10*	2.22E-07	-
				rs936551	*CPLX1*	5.29E-08	-
				rs11248047[b]	*CPLX1*	1.26E-07	Not associated in Turkish, MEA, Greek, UK, Korean and Japanese [Remmers *et al*., 2010]
				rs2848713	*HLA-B*	3.99E-45	-
				rs9260997	*HLA-A*	5.49E-09 *	-
Hou *et al*., 2012	149 vs 951	Chinese	906,600 SNPs	rs6692084	*DEPDC1*	2.81E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs12134670	*DEPDC1*	3.13E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs1472224	*DTL*	5.73E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs1465825	*DNMT3A*	3.38E-07	Not associated in chinese [Hou *et al*., 2012a]
				rs17006292	*TFCP2L1*	1.03E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs6744214	*PSMD14*	1.67E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs6733456	*PSMD14*	1.98E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs2390639	*STK39*	3.97E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs3769393	*STK39*	6.17E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs897200	*STAT4*	5.88E-05	Replicated in chinese [Hou *et al*., 2012a]
				rs7574070	*STAT4*	8.56E-05	Replicated in chinese [Hou *et al*., 2012a] and in Turkish and Japanese [Kirino *et al*., 2013]
				rs7572482	*STAT4*	9.77E-05	Replicated in chinese [Hou *et al*., 2012a]
				rs17562982	*SGPP2*	1.91E-09	Not associated in chinese [Hou *et al*., 2012a]
				rs7561555	*ASB18*	4.70E-08	Not associated in chinese [Hou *et al*., 2012a]
				rs13435197	*SLIT2*	3.59E-08	Not associated in chinese [Hou *et al*., 2012a]
				rs4493590	*SORBS2*	4.88E-06	Not associated in chinese [Hou *et al*., 2012a]
				rs10516130	*MSX2*	2.98E-06	Not associated in chinese [Hou *et al*., 2012a]
				rs4959053	*HLA-B*	2.29E-20	-
				rs12194547	*C6orf85*	1.91E-05	Not associated in chinese [Hou *et al*., 2012a]
				rs2190411	*ABCB5*	8.77E-10	Not associated in chinese [Hou *et al*., 2012a]

				rs27829332	*SUSD1*	2.47E-09	Not associated in chinese [Hou *et al.*, 2012a]
				rs420798	*API5*	1.79E-05	Not associated in chinese [Hou *et al.*, 2012a]
				rs16937370	*API5*	6.01E-09	Not associated in chinese [Hou *et al.*, 2012a]
				rs549630	*SLC43A3*	2.04E-08	Not associated in chinese [Hou *et al.*, 2012a]
				rs2895135	*RIMBP2*	3.35E-09	Not associated in chinese [Hou *et al.*, 2012a]
				rs12589991	*GALNTL1*	2.16E-08	Not associated in chinese [Hou *et al.*, 2012a]
				rs749240	*SMG6*	6.43E-09	Not associated in chinese [Hou *et al.*, 2012a]
				rs798887	*LILRB1*	2.23E-05	Not associated in chinese [Hou *et al.*, 2012a]
				rs103294	*LILRA1*	2.19E-05	Not associated in chinese [Hou *et al.*, 2012a]
				rs6082210	*C20orf74*	7.01E-06	Not associated in chinese [Hou *et al.*, 2012a]
				rs817277	*CDH26*	3.24E-05	Not associated in chinese [Hou *et al.*, 2012a]
				rs817283	*CDH26*	6.42E-05	Not associated in chinese [Hou *et al.*, 2012a]
Lee *et al.*, 2013	379 vs 800	Korean	906,600 SNPs	rs17275413	*C3orf58*	4.49E-06	-
				rs2293595	*YTHDC1*	4.61E-06	-
				rs10033058	*YTHDC1*	2.38E-06	-
				rs17089267	*YTHDC1*	3.32E-06	-
				rs1965673	*WWC1*	4.91E-07	-
				rs1465400	*WWC1*	1.58E-06	-
				rs4947296	*HLA-B*	4.01E-13	-
				rs11769828	*GIMAP4*	1.60E-06	-
				rs1916012[a]	*GIMAP4*	6.19E-06	Replicated in Japanese [Lee *et al.*, 2013]
				rs1522596[a]	*GIMAP4*	7.70E-06	Replicated in Japanese [Lee *et al.*, 2013]
				rs1608157[a]	*GIMAP4*	2.83E-06	Replicated in Japanese [Lee *et al.*, 2013]
				rs10266069[a]	*GIMAP2*	2.67E-04	Replicated in Japanese [Lee *et al.*, 2013]
				rs10256482[a]	*GIMAP2*	5.27E-04	Replicated in Japanese [Lee *et al.*, 2013]
				rs2286900[a]	*GIMAP1*	3.53E-05	Not associated in Japanese [Lee *et al.*, 2013]
				rs4239774	*PLCB1*	2.07E-06	-

1.1.9.4 Candidate genes association studies

In addition to the *HLA-B* locus, several candidate genes such as intercellular adhesion molecule-1 (*ICAM1*) [Verity *et al*., 2000; Kim *et al*., 2003], endothelial nitric oxide synthase (*eNOS*) [Salvarani *et al*., 2002; Karasneh *et al*., 2005b; Oksel *et al*., 2006], glutathione S-transferase (*GST*) [Uzunoğlu *et al*., 2006], and vascular endothelial growth factor (*VEGF*) [Nam *et al*., 2005] genes have been tested and found associated to BD susceptibility. However, most of these studies were conducted on a limited number of patients, the significance levels were weak, the positive findings have not been replicated, and some results are conflicting. Two exceptions were *IL10* and *IL23R* which were found associated with BD by a candidate gene approach [Wallace *et al*., 2007; Jiang *et al*., 2010] previously to their identification as susceptibility genes for BD in two GWAS [Remmers *et al*., 2010; Mizuki *et al*., 2010]. In the paper by Remmers *et al*., (2010), they also confirmed the association of BD with genes previously reported as associated such as the Fc receptor-like protein 3 (*FCRL3)* [Li *et al*., 2008], the cytotoxic T-lymphocyte antigen 4 [Gunesacar *et al*., 2007], the solute carrier family 11 member 1 (*SLC11A1)* [Ateş *et al*., 2009; Kim *et al*., 2006], the *IL17F* [Jang *et al*., 2008], the toll-like receptor 4 (*TLR4)* [Meguro *et al*., 2008; Horie *et al*., 2009] and the caspase recruitment domain-containing protein 15 (*CARD15)* [Ahmad *et al*., 2005] genes, although these associations were not among the top GWAS findings [Remmers *et al*., 2010].

1.1.10 Management

Behçet's disease is characterized by relapses and remissions, and highly variable clinical course, even though the frequency and severity of its symptoms may diminish with time. After the fourth decade, the clinical severity reduces, with clinical complications generally recurring at longer intervals [Bardak, 1999]. The treatment of BD is symptomatic and empirical, lacking specific disease therapy and being generally specific to the clinical features of each patient. Treatment is usually multidisciplinary, requiring close collaboration among specialists in oral medicine, dermatology, ophthalmology, and others. Male patients and those with early-onset disease usually require more aggressive treatment than do other affected individuals [Yazici *et al*., 1999].

1.1.10.1 Therapeutics

Corticosteroids (systemic or topic) are commonly used to treat clinical BD manifestations in the form of monotherapy or in combination with immunosuppressants (such as Azathioprine, Ciclosporine A or Infliximab). Although they have beneficial effects against acute inflammation, studies have shown that there are still persistent risks for morbidity, mainly for ocular and neurological compromised patients, in addition to the frequent adverse reactions. Therefore, new therapeutic modalities, such as those with biological agents have been introduced on the market and constituted an alternative BD treatment option [Cho *et al*., 2012; Tunes *et al*., 2009].

The earliest biological response modifier introduced were the interferons, such as, IFN-α-2a that are cytokines with antiviral, antitumoral and immunoregulatory activity which appear to be effective in the BD treatment because of their action of inhibiting T lymphocytes and stimulating NK cells [Tunes *et al*., 2009]. A multitude of reports have described its benefits for the uveitis associated with BD and many patients enjoy durable remissions of their ocular inflammatory disease even after discontinuation of therapy. Its action was also effective in the remission of mucocutaneous manifestations. However side-effects are almost universal and some can be dangerous [Tunes *et al*., 2009; Benitah *et al*., 2011; Yalçindağ *et al*., 2012].

Of the newer biological response modifiers TNF blocking drugs such as infliximab (chimeric anti-TNF-α monoclonal antibody), etanercept (fusion protein human p75 TNF-α receptor IgG1), and adalimumab (humanized anti-TNF-α monoclonal antibody) have been reported to have some success in patients with Behçet's disease who have mucocutaneous, gastrointestinal lesions and also neurological manifestations [Arida *et al*., 2011]. There is enough published experience to suggest that TNF blockade represents an important therapeutic advance for patients with severe disease who are resistant to standard immunosuppressive regimens and for those patients with contraindications or intolerance to these treatments and therefore its use is increasing [Arida *et al*., 2011]. It was found, *in vivo* and *in vitro* that infliximab is able to suppress in γδ T cells expansion, activation and cytotoxic activity [Accardo-Palumbo *et al*., 2010]. At present, however, the high cost, need for injections, troublesome toxic side effects, and lack of long term evidence for therapeutic efficacy are limiting the widespread acceptance of anti-TNF-α agents as a first-line choice for BD management [Cho *et al*., 2012].

1.1.10.2 Prognosis

The treatment of BD remains unsatisfactory due to the rarity and heterogeneity of the condition, uncertainties about the relevant aetiology and pathogenesis, and the lack of data from well-controlled clinical trials. In particular, the management of severe manifestations, such as vascular and neurological involvement, remains entirely empirical [Marshall, 2004]. Combination of therapies is generally accepted to be more effective than single agent regimens. The most effective management involves early diagnosis and clinical intervention with continuous follow-up.

The age of onset seems to be a prognostic factor, with patients with early onset (before 25 years of age) having a poorer prognosis than those with onset after 40 years of age [Yazici *et al*., 1984]. Nevertheless, there is considerable variation among patients. While BD runs an indolent course in some patients, with mucocutaneous manifestations dominating the clinical picture, others have serious complications, such as ocular and neurological involvement that may result in significant disability and can greatly diminish quality of life [Koç *et al*., 1992]. Mortality was found to occur in 9.8% of BD patients [Kural-Seyahi *et al*., 2003], mainly due to major vessel disease, neurological involvement, and perforated intestinal ulcers [Yazici *et al*., 1996; Kural-Seyahi *et al*., 2003].

Behçet's disease thus remains a significant challenge to clinicians of many specialties. However, in view of the new findings made for BD at the genetic and immunological level it should now be possible to investigate therapies for this potentially serious disorder with greater focus than was previously possible [Marshall, 2004].

1.2 GENOMIC APPROACHES FOR THE STUDY OF COMPLEX DISEASES

1.2.1 Family linkage studies

Linkage analysis can be used to identify regions of the genome that contain genes that predispose to disease. The methods used to search for linkage in multifactorial diseases are model based (i.e., parametric) and model free (i.e., non-parametric). Parametric linkage analysis calls for the researcher to specify a model of inheritance and estimate the frequency and penetrance of the disease genes, and it is a powerful approach when applied to extended pedigrees with many affected individuals. Non-parametric linkage analysis is less powerful but does not require the specification of a genetic model and instead relies on the principle that affected relatives share haplotypes that are identical by descent (IBD) in the region of a disease causing gene more often that would be expected by chance [Visscher *et al*., 2012; Juran *et al*., 2007]. The test used to calculate the significance of the observations indicating that two loci are linked at a given recombination fraction (ranging from $\vartheta = 0$ to $\vartheta = 0.5$) is the LOD score test. It evaluates the overall likelihood of the pedigree, on the alternative assumption that the loci are linked ($\vartheta < 0.5$) or not linked ($\vartheta = 0.5$). The ratio of these two likelihoods gives the odds of linkage and it is expressed as the log of this ratio (LOD stands for "log of odds") [Giordano, 2005].

The alleles from the polymorphic markers will segregate together with causal variants within the pedigree and the meiotic recombination within the families can be explored to narrow the chromosome region to search for the susceptibility gene. The resolution of the linkage mapping is limited by the fact that relatively few meiotic breaks occur over the small number of generations available in most pedigrees and therefore linkage studies typically provide target intervals of several centimorgans or megabases. However, functional annotation of the genes in the linkage region makes it easier to select likely susceptibility determinants within broad regions. The markers of choice for linkage studies are the microsatellites (for example di- and tetranucleotide repeats) because they are often vey polymorphic and therefore informative and suitable for recovering the maximum information from a pedigree [Visscher *et al*., 2012].

Linkage mapping has been extremely successful in identifying highly penetrant, rare alleles involved in Mendelian diseases (e.g., single gene disorders) however mapping loci underlying common disorders has not been as successful [Visscher *et al*., 2012]. Clinical, locus, and allelic heterogeneities are common features of complex diseases that dilute the

linkage signal, reducing significantly the chance of identifying plausible candidate regions. Moreover, parametric approaches are impractical, since models of inheritance and estimations of the frequency and penetrance of the disease genes are not readily assumable for complex disorders. In addition, the late age of onset of most complex diseases often prevents the assessment of multiple generations, significantly reducing the power of both parametric and nonparametric approaches. However, the use of linkage analysis to identify genes involved in the development of complex disease could prove useful, especially when applied to families in whom the genetic component is likely to be enriched, such as those with an unusually high rate of disease occurrence or exceptionally early age of disease onset. Although these families might not be representative of the disease in the majority of the population, findings could implicate major networks or pathways involved in the disease and would certainly provide a basis for further investigations [Juran *et al*., 2007].

1.2.2 Association studies

Association studies aim to correlate differences in allele frequencies at a polymorphic marker between groups of individuals with a different trait. The trait may be some quantitative characteristic or a discrete attribute or disease (e.g., cases and controls). Typically, association studies involve case-control population samples but family-based association tests using for example trios of parents and an affected child are also useful.

Most association studies use panels of SNPs as markers. There are 5 important advantages of using SNPs rather than other types of genetic polymorphism to investigate the genetic determinants of complex human diseases: first, SNPs are frequent throughout the genome, being found in exons, introns, promoters, enhancers and intergenic regions, and some of these polymorphisms might themselves be functional; second, groups of adjacent SNPs might exhibit patterns of correlations that could be used to enhance gene mapping and which may highlight recombination hot-spots; third, interpopulation differences in SNP frequencies can be used in population-based genetic studies (e.g. to control for population stratification); and fourth, SNPs are less mutable than other types of polymorphisms and this greater stability could allow more consistent estimates of genotype-phenotype associations [Palmer *et al*., 2005]. SNPs are also typically bi-allelic which simplifies analysis and makes them suitable for high-throughput automated array "chip" genotyping technologies [Collins, 2009].

1.2.2.1 Basic principles

Association can be tested in a direct or indirect form. Direct association target polymorphisms which are themselves putative causal variants. This type of study is the easiest to analyse and the most powerful, but the difficulty is the selection of candidate gene polymorphisms, since there is not enough knowledge to predict which variants will cause variation in gene regulation and expression or differential splicing. Non-synonymous coding variants – those that alter the amino-acid sequence in the gene product – are obvious targets, but assessment of the potential regulatory effect of intronic variants or those lying several kilobases (kb) upstream of a gene remain poor [Hattersley *et al*., 2005]. Thus direct association only has the potential to discover some of the genetic causes of disease and disease-related traits [Cordell *et al*., 2005].

The mapping of susceptibility genes for complex disorders by the indirect method is the most widely used in complex diseases and depends on the existence of LD, at a population level, between the causal variants and nearby markers [Cordell *et al*., 2005]. LD is different from family linkage because it depends on meiosis accumulated over many generations, at the population level. When LD is high, the redundancy between markers implies that most of the genetic diversity can be captured genotyping few markers, which are called haplotype tagging SNPs (htSNPs). Linkage disequilibrium is also important for the haplotype blocks. Genetic loci across large areas of the genome were suggested to divide into blocks characterized by little disequilibrium between blocks and limited haplotype diversity within blocks [Cordell *et al*., 2005]. Defining and genotyping a relatively small number of these htSNPs could allow unambiguous determination of the common haplotypes in a population and capture all or most of the LD within that region. By this means, SNP-phenotype association studies can be done relatively efficiently, by contrast with genotyping all common variants in a given genomic region or in the entire genome [Palmer *et al*., 2005]. The international HapMap project was aimed at constructing genome-wide maps of LD patterns in multiple populations. The project genotyped at a density of more than 1 SNP per 1kb in samples collected in multiple continents. This allowed the construction of high density maps and to define a list of htSNPs that capture most of the common genomic variation in a number of human populations, clarifying the value of each SNP for the indirect association mapping of disease genes [International HapMap Consortium, 2003].

LD will remain crucial to the design of association studies until whole-genome resequencing becomes routinely available. Currently, few of the more than 10 million

common human polymorphisms are typed in any given study. If a causal polymorphism is not genotyped, we can still hope to detect its effects through LD with polymorphisms that are typed. To assess the power of a study design to achieve this, we need to measure LD. However, LD is a non-quantitative phenomenon: there is no natural scale for measuring it. A wide variety of statistics have been proposed to quantify the amount of LD, and these have different strengths, depending on the context. Most of the measures of LD that are in wide use quantify the degree of correlation between pairs of markers. In part, they differ according to the way in which they depend on the marginal allele frequency [Pritchard *et al*., 2001]. Considering two loci, with minor alleles *a* and *b* and major alleles *A* and *B*, if the frequencies of the alleles at these two loci are *pA*, *pa*, *pB* and *pb* in a population of inference, then one measure of the magnitude of LD is defined by [Vens *et al*., 2012] :

$$D = pAB - pA.pB$$ (Please refer to Table 4)

Whether *D* is positive or negative depends on the arbitrary labelling of alleles. The maximum value that *D* can have depends strongly on allele frequency: *Dmax* is the lesser of *pApb* or *papB* if *D* is positive or *pApB* or *papa* if *D* is negative. A measure that attempts to avoid this dependence on allele frequency is estimation of D', defined by [Vens *et al*., 2012]:

$$D' = \frac{D}{D\max}$$

The measure D' ranges, by definition, between -1 and 1, and its absolute value is equal to 1 whenever one entry on the Table 4 is equal to 0 (i.e., at most three haplotypes observed), a situation that arises if no recombination has occurred between the loci since the more recent founding mutation. Because the sign is arbitrary, $|D'|$ is often used rather than D' [Vens *et al*., 2012].

Table 4. 2x2 contingency table of haplotype possibilities for two markers. *A* and *a* represent the alleles of marker 1 and *B* and *b* the alleles of marker 2.Table adapted from Vens et al., 2012.

Alelles	***B***	***b***
A	*pAB*	*pAb*
a	*paB*	*pab*

Among other popular measures of LD is the square of the correlation coefficient between the A and B loci, r^2, defined by:

$$r^2 = \frac{D^2}{pA.pa.pB.pb}$$

The definition of r^2 can be understood by considering the alleles as realizations of quantitative random variables (with values 0 and 1), among which we calculate a correlation coefficient. The measure r^2 ranges between 0 and 1, and it is equal to 1 only when two entries of the table 4 are equal to 0 [Chen *et al.*, 2006]. When $r^2=1$ the SNPs are in perfect LD, meaning that the two SNPs have not been separated by recombination, and also have the same allele frequencies.

If a SNP has a low minor allele frequency (MAF), it is quite possible that the rare haplotype that carries it, is not observed in a small sample. This leads to a D' being equal to 1, irrespective of the level of linkage disequilibrium [Chen *et al.*, 2006]. Hence, the range of values that D' can take is not dependent on the allele frequencies, and it may more accurately reflect the pattern of recombination. There is a large literature discussing the choice of these measures. Typically, r^2 is preferred when the focus is on the predictability of one polymorphism given the other (and hence it is often used in power studies for association designs). D', instead, is the measure of choice to assess recombination patterns (haplotypes blocks have often been defined on the basis of D'). Despite their effectiveness, these measures suffer from two limitations: (a) they are not easily generalizable to multiallelic markers; (b) they are defined on the population haplotype distribution, and their performance can be rather unsatisfactory when applied to the empirical distribution derived from a finite sample [Chen *et al.*, 2006].

Both D' and r^2 are two-locus measures; however, with dense markers it is of interest to summarize LD over a region. One approach is to compute local averages of pairwise values of D' and r^2. Alternatively, values over a region can be illustrated diagrammatically with colours encoding different values (LD plots) [Balding, 2006].

1.2.2.2 Type of studies

Candidate gene approach

Evidence from a range of sources (e.g., biology, pharmacology, animal models, prior association data, etc) can be used to identify the so-called candidate genes with higher likelihood for phenotypic involvement, and before the GWAS era almost all association studies featured genes selected based on these criteria. However, poor understanding of the molecular mechanisms underlying most complex traits (itself one of the main justifications for gene discovery efforts) turns imprecise the calculation of the prior odds associated with any given gene [Hattersley *et al*., 2005].

Genome-wide association studies

In contrast to methods which specifically test one or a few genetic regions, GWASs use high-throughput genotyping technologies to assay hundreds of thousands of genetic variants, usually SNPs, in thousands of individuals and therefore represents a powerful tool for investigating the genetic architecture of complex diseases.

Many of the genes found associated in the GWAS were not previously suspected of having a role in the disease under study, and some are located in genomic regions containing no known genes. Although GWASs are unbiased with respect to prior biological knowledge (or prior beliefs) and with respect to genome location, they are not unbiased in terms of what is detectable. GWASs rely on LD between genotyped SNPs and ungenotyped causal variants and since the SNPs that are on the commercial arrays have been selected to be common (most having a minor allele frequency (MAF) >0.05), GWASs are by design powered to detect association with causal variants that are relatively common in the population (Figure 4) [Visscher *et al*., 2012].

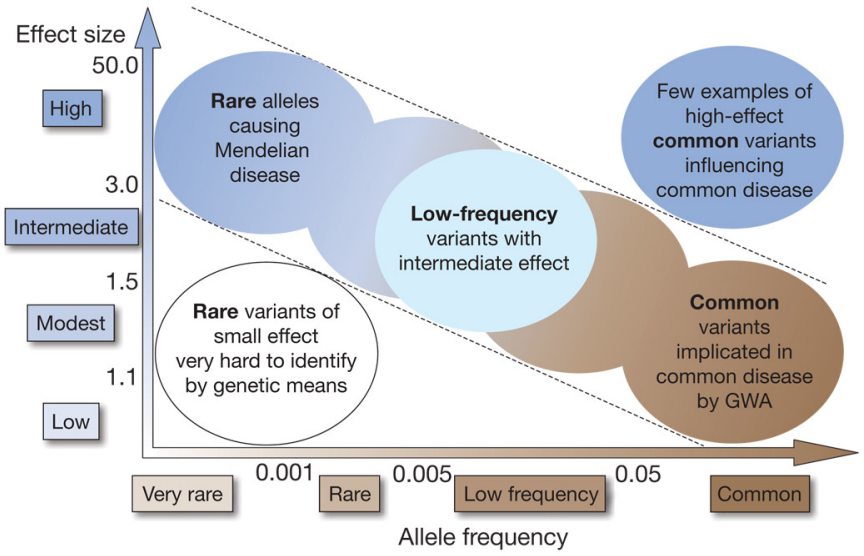


Figure 4. Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Figure taken from Manolio *et al.*, 2009.

Furthermore, the SNPs included in arrays vary between being haplotype tagging SNPs to random SNPs throughout the genome, and they do not cover all the variation existing in the genome.

Although these studies are clearly many steps removed from actual clinical use, GWAS can facilitate new unbiased biologic insights into disease pathogenesis, and also provide more definite answers for the cause of the diseases of complex genetic traits when performed in large groups and confirmed in different populations [Pearson *et al.*, 2008].

Next generation sequencing

The primary technology for the detection of new and rare variants is sequencing. Exhaustive sequencing of genomic regions has recently become feasible for the first time with the advent of next generation sequencing (NGS) technologies that process millions of sequence reads in parallel. Given the lack of good representation of SNPs with a MAF smaller than 5% in current GWAS arrays, a comprehensive catalogue of SNPs with a prevalence of 1 to 5% is being generated by the 1000 genomes project (1000 Genomes: a deep catalogue of human genetic variation, http://www. 1000genomes.org/page.php). Resequencing of an associated gene/region may identify additional, previously unknown

variants (frequency <1%) with a possible biological role [Manolio, 2010]. Once the variant is sequenced, genotyping can be performed in bigger sample sizes at a lower cost.

1.2.2.3 Important concerns in genetic association studies

Genetic association studies of complex phenotypes often either failed to discover susceptibility loci or failed to replicate studies that did. Despite the widespread use of genetic case-control studies, their inconsistency is a generally recognized limitation. This lack of reproducibility is often ascribed to small sample size with inadequate statistical power, biological and phenotypic complexity, population-specific LD, effect-size bias and population stratification. Other reasons for the non-replication of true positive association results include differences in the study design and analytical method, inter-population heterogeneity, phenotype definition, environmental exposures not considered and markers genotyped [reviewed in Palmer *et al*., 2005]. It is now routinely argued that large sample sizes (generally thousands rather than hundreds), rigorous *P*-value thresholds and replication in multiple independent datasets are necessary for reliable results [Palmer *et al*., 2005]. Power for studies of allelic association will depend also on the effect size of the susceptibility locus, the strength of LD with a marker, and the frequencies of susceptibility and marker alleles [Palmer *et al*., 2005].

Defining a phenotype

The careful characterization of an appropriate phenotype is extremely important in the search for susceptibility genes [Collins, 2009]. In many association studies a qualitative (case versus control) phenotype is used. It is essential that phenotypes are strictly defined to achieve maximum homogeneity and facilitate future replication studies [Collins, 2009]. Controls should be selected from the same population as cases, or methods must be used to correct for any latent population stratification.

Power of sample size

The key determinant of quality in an association study is sample size. With the remote chance of finding common genes with large effects, studies must be powered to detect variants that are common but have low relative risk, or rarer but with higher relative

risk, which means sample sizes of thousands [Hattersley *et al*., 2005]. Overall, association studies are capable of identifying substantial genetic effects (i.e., OR > 2.0) on disease phenotype with relatively small sample sizes (n~200) [Whitcomb *et al*., 2005] and have high power to detect small effects of genetic variation (i.e., OR < 2.0) but require the sample sizes to be quite large (n~1000) [Juran *et al*., 2007]. Rare variants with low relative risks are largely beyond the reach of genetic epidemiology because of the massive sample size that would be needed (Figure 4) [Hattersley *et al*., 2005].

Population stratification

The presence of population stratification – allele frequency differences between cases and controls due to ancestry differences - can mimic the signal of association and lead to false positives results or to miss real effects (Figure 5). Such stratification may arise from recent admixture or from poorly matched cases and controls.

The most obvious way of avoiding this difficulty is to measure association in well-mixed, outbred populations. Failing this, one method of dealing with this problem is to match cases and controls (in the design and/or the analysis) by geographical region and by any markers of ethnic origin [Cordell *et al*., 2005]. The ancestries can be effectively discerned using dense genotyping data, as in the case of the GWAS, making it possible to correct for population stratification and to identify ancestry-specific risk loci. Principal component analyses or multidimensional scaling methods are commonly used to identify and remove individuals exhibiting divergent ancestry before association testing. To adjust for any residual population structure during association testing, the principal components from principal component analyses (PCA) or multidimensional scaling methods can be included as covariates in a logistic regression [Clarke *et al*., 2011] (for a brief introduction to the logistic regression methods please refer to the chapter 1.2.2.4, page 41). However in target association studies, such as candidate gene studies or replication studies, the typical number of analyzed markers is much smaller making it far more difficult or impossible to infer ancestry. In these situations, a method for dealing with population stratification is to genotype ancestry informative markers (AIMs) – loci with allele frequencies that differ between the founder populations - and use them to correct for stratification in association studies [reviewed in Cordell *et al*., 2005].

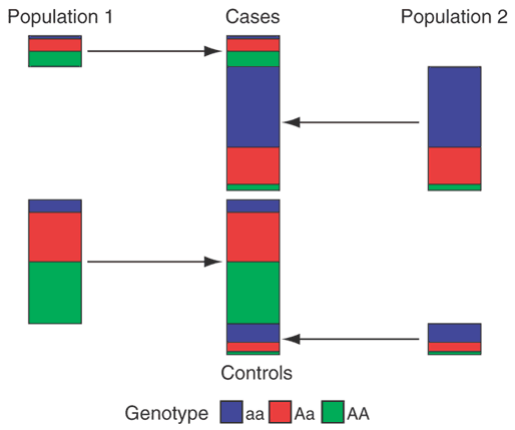


Figure 5. The effects of population structure at a SNP locus. The figure shows two populations 1 and 2, in which the cases have an excess of individuals from population 2 and population 2 has a lower frequency of allele A than population 1. In this example, the structure mimics the signal of association in that there is a significant difference in allele and genotype frequencies between cases and controls. Figure taken from Marchini *et al*., 2004.

Multiple testing

Multiple testing is an issue in many genetic association studies where multiple SNPs per gene or multiple SNPs in several loci are tested. This problem is exacerbated in GWAS where over a million of SNPs are tested and there is a great potential for false positive results. Assuming the independence of SNPs, selecting 0.05 as the level of significance, and using a million markers, we expect to obtain 50,000 false-positive associations due to chance alone. Thus there is a need for some form of statistical correction. It has become conventional to apply the Bonferroni correction, in which the conventional *P*-value of 0.05 is divided by the number of tests performed. In the case of testing one million SNPs, this correction implicates using a threshold of 5.00E-08 that is now accepted as the *P*-value for a SNP to have genome-wide significance in the discovery population [Roberts *et al*., 2010]. However this correction has been criticized as being too conservative since it does not proper account for the correlation between SNPs in LD. There are several statistical techniques to correct for multiple comparisons, but replication of genetic association findings in independent datasets remains the gold standard for complex disease genetics.

Understand genotype-phenotype relationship

There is a substantial gap between SNP associations from a GWAS and understanding how the locus contributes to disease. Further genotyping and statistical analyses are often necessary to identify causal variants, which are then functionally investigated. In general, follow-up functional studies aimed at explaining the disease mechanism underlying any detected associations would be beneficial. However, such studies are often not performed because of lack of a suitable model system and/or adequate specimens from which to derive RNA or protein. Moreover, when they are performed, the results can be unsatisfying and difficult to interpret, as the contextual milieu in which function is affected and contributes to disease is likely to be altered or lost [Juran *et al*., 2007].

1.2.2.4 Analysis of association studies

Preliminary analyses

Data quality is of paramount importance, and data should be checked thoroughly, for genotyping errors (e.g., using samples with known genotypes and/or sample duplication), batch or study-centre effects, sample missing calls, or for unusual patterns of missing data. Testing for Hardy–Weinberg equilibrium (HWE) can also be helpful. Deviations from HWE can occur due to inbreeding, population stratification or selection. So far, researchers have tested for HWE primarily as a data quality check and have discarded loci that, for example, deviate from HWE among controls at a certain significance level (e.g., *P*-value= 1.00E-03 or 1.00E-04). However, apparent deviations from HWE can arise in the presence of a common deletion polymorphism, because of a mutant polymerase chain reaction (PCR)-primer site or because of a tendency to miscall heterozygotes as homozygotes. Therefore, the possibility that a deviation from HWE is due to a deletion polymorphism or a segmental duplication that could be important in disease should be considered before discarding loci. Testing for deviations from HWE can be carried out using a Pearson goodness-of-fit test, often known simply as 'the χ^2 test' because the test statistic has approximately a χ^2 null distribution. However, there are many different χ^2 tests. The Pearson test is easy to compute, but the χ^2 approximation can be poor when there are low genotype counts, and it can be a best choice to use a Fisher exact test, which does not rely on the χ^2 approximation [Balding, 2006].

Models and measures of association

Considering a SNP with a minor allele *a* and a major allele *A*, the possible genotypes will be *a/a*, *A/a* and *A/A*. The disease penetrance associated with a given genotype is the risk of disease in individuals carrying that genotype. Standard models for disease penetrance that imply a specific relationship between genotype and phenotype include additive, common recessive and common dominant models. Assuming a genetic penetrance parameter γ (γ > 1), an additive model indicates that risk of disease is increased γ-fold for genotype *a/A* and by 2γ-fold for genotype *A/A*; a common recessive model indicates that two copies of allele *A* are required for a γ-fold increase in disease risk, and a common dominant model indicates that either one or two copies of allele *A* are required for a γ-fold increase in disease risk [Clarke *et al*., 2011].

Association testing

Tests of association are usually performed separately for each individual SNP. The data from a SNP with a minor allele *a* and a major allele *A* can be represented as a contingency table of counts of disease status (cases versus controls) by either allele count (e.g., *a* and *A*) or genotype count (e.g., *a/a*, *A/a* and *A/A*) (see Table 5).

Table 5. 2x2 and 2x3 contingency table of disease status by allele count and by genotype count, respectively. *a* represents the minor and *A* the major allele. Table adapted from Clarke *et al*., 2011.

Alelles	**a**	**A**	**Total**
Cases	m11	m12	m1.
Controls	m21	m22	m2.
Total	m.1	m.2	2n

Genotype	**a/a**	**a/A**	**A/A**	**Total**
Cases	n11	n12	n13	n1.
Controls	n21	n22	n23	n2.
Total	n.1	n.2	n.3	n

Under the null hypothesis of no association with the disease, we expect the relative allele or genotype frequencies to be the same in case and control groups. Pearson's chi-squared test (χ^2) and Fisher exact test can be used to test for independence of the rows and columns of the contingency table. In complex traits, it is widely believed that contributions to disease risk from individual SNPs are often roughly additive — that is, the heterozygote risk will be intermediate between the two homozygote risks. Therefore, the allelic association test

with 1 degree of freedom (d.f.) will be more powerful than the genotypic test with 2 d.f., if the genotype risk is additive. However, in the absence of HWE in controls, the allelic association test is not suitable and alternative methods must be used to detect additive risks [Clarke *et al*., 2011]. One alternative is to adopt the 1 d.f. Cochran-Armitage trend Test. The Cochran-Armitage test is more conservative than the χ^2 test and does not rely on an assumption of HWE. The idea is to test the hypothesis of zero slope for a line that fits the three genotypic risk estimates best (Figure 6). The Cochran-Armitage trend test, in which each copy of the allele is assumed to increase risk by the same amount (additive model), is the most common test used in GWASs[Clarke *et al*., 2011; Balding, 2006]. The test has good power in this case but power is reduced by deviations from additivity. In an extreme scenario, if the two homozygotes have the same risk but the heterozygote risk is different (overdominance), then the Armitage test will have no power for any sample size even though there is a true association [Balding, 2006].

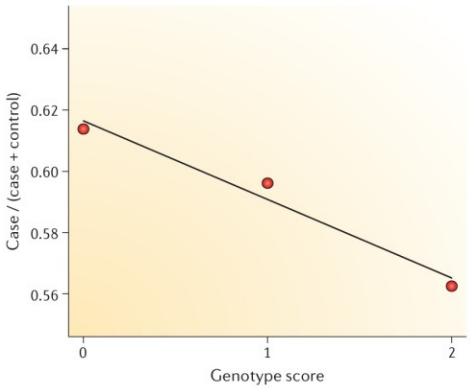


Figure 6. Armitage test of single-SNP association with case–control outcome. The dots indicate the proportion of cases, among cases and controls combined, at each of three SNP genotypes (coded as 0, 1 and 2), together with their least-squares line. The Armitage test corresponds to testing the hypothesis that the line has zero slope. Here, the line fits the data reasonably well as the heterozygote risk estimate is intermediate between the two homozygote risk estimates; this corresponds to additive genotype risks. Figure taken from Balding, 2006.

Logistic regression models are used when there is a need to include additional covariates that we expect to modify disease risk. These covariates may include demographic variables (such as sex and age), environmental factors (e.g., smoking habits), population stratification (e.g., principal component capturing variation due to differential ancestry), and also other genetic variants. In logistic regression models, the logarithm of the odds of disease

is the response variable, with linear (additive) combinations of the explanatory variables (genotype variables and any covariates) entering into the model as its predictors [Clarke *et al*., 2011].

In the case of association studies performed using the same markers (e.g., SNPs), their results can be combined in meta-analysis. Meta-analysis can increase power, identify false-positives due to biases in individual studies and highlight heterogeneity. The heterogeneity between studies can be evaluated using Cochrane's Q and I^2 statistics that should be used combined. Generally, if there is no heterogeneity, a fixed-effect model which assumes that a single common effect underlies every study in the meta-analysis will be applied. A random effect model should be considered when heterogeneity is present [Roberts *et al*., 2010].

Another popular strategy in association studies, suggested by the blocklike structure of the human genome, is to perform haplotype analysis of SNPs in a chromosomal region of interest. This approach has greater power than single SNP analysis as it can capture the combined effects of tightly linked causal variants and can lead to analyses with fewer degrees of freedom. Perhaps more importantly, haplotype-based association testing may add additional information for detecting causal variants beyond that contained in the analysis of any of the single SNP that make up the observed haplotypes. But this benefit is minimized when SNPs are ascertained through a tagging strategy. A standard approach is to compare haplotype frequencies between cases and controls based on the estimated haplotype frequencies (since the haplotypes are not observed). If genome-wide SNP data are available, then either a block-based or sliding window-based method can be considered to look for contiguous SNPs. The block-based approach of Gabriel *et al*., (2012) [Gabriel *et al*., 2002], is the most common method used to visualize LD blocks, but has an important drawback when it comes to defining groups of SNPs to be used in haplotype association: many SNPs are not included in any block by the algorithm and their association in haplotypes cannot be assessed [Stram *et al*. 2012]. After haplotype assignments, the simplest analysis involves testing for independence of rows and columns in a 2×k contingency table, where k denotes the number of distinct haplotypes [Balding, 2006].

In complex diseases, gene-gene interactions should also be considered. This type of interaction is difficult to detect and characterize using traditional parametric statistical methods such as logistic regression. The multifactor dimensionality reduction method (MDR) was developed as a non-parametric and genetic model-free data mining strategy for identifying combination of SNPs that are predictive of a discrete phenotype. This software collapses high-dimensional genetic data into a single dimension by pooling multilocus

genotypes into high- and low-risk groups. The new multilocus genotype attribute is then tested for its ability to classify and predict disease status. False-positive results due to multiple testing are reduced through combination of the cross-validation strategy and permutation testing. Entropy-based interaction dendrograms are used for interpreting epistasis models [Ritchie *et al*, 2001; Moore, 2006].

Relative risk and odds ratio

The strength of the association between a marker and a disease is usually described by the relative risk (RR) or the OR. These parameters are widely used in epidemiology to assess the risk conferred by the exposition to a risk factor that in genetics is the presence or the absence of a certain marker allele at a specific locus disease [Clarke *et al*., 2011].

The allelic OR describes the association between disease and allele by comparing the odds of disease in an individual carrying allele *A* to the odds of disease in an individual carrying allele *a*. The allelic odds ratio is estimated by [Clarke *et al*., 2011]:

$$OR_A = \frac{m12m21}{m11m22}$$ (Please refer to Table 5 of this chapter, page 40)

RR estimates based on penetrance can only be derived directly from a prospective cohort study, in which a group of individuals from the same population exposed and unexposed to certain allele are followed up to assess who develops disease [Clarke *et al*., 2011]. If the disease prevalence in a group of individuals carrying an *a* allele can be estimated and is denoted *P0*, then the relative risk of disease in individuals with an *A* allele compared with an *a* allele is estimated by [Clarke *et al*., 2011]:

$$RR_A = \frac{OR_A}{1 - P_0 + P_0 OR_A}$$

Imputation

The method of inferring genotypes at SNPs not effectively genotyped given the known LD pattern among SNPs within haplotypes is called imputation. This technique has greatly improved coverage of the commercial arrays, thereby increasing the power of

genome-wide association studies and allowing genotypes obtained on different commercial arrays to be compared. However, despite its remarkable utility in augmenting the power of GWAS at no additional cost, there are important limitations to imputation. For example, for populations with complex haplotype structures imputation is of little utility. In addition, even small rates of genotyping error or missing data (as little as 5%) markedly reduce the accuracy of imputation and low frequency SNPs (MAF lower than 5%) cannot be accurately imputed [Huang *et al*., 2009]. Therefore, associations of many rare SNPs with stronger effects than common SNPs will not be revealed by imputation [Roberts *et al*., 2010].

1.2.3 Gene expression studies using microarrays

Alterations from normal physiology are frequently accompanied by biochemical changes, resulting in changes in gene expression patterns. It has become clear that complex diseases result from collective actions of many genetic and nongenetic factors. Therefore, genetic dissection of complex diseases should be carried out in a global context. The patterns of upregulation or downregulation of gene activities can serve as secondary endpoints or biomarkers for disease [Gu *et al*., 2002]. Furthermore, the assumption of no a priori etiological hypotheses provides an unbiased approach to elucidating the pathogenesis of complex diseases. Global gene expression analysis uses microarrays, sequencing, and other methods to measure the levels of RNA species in biological systems.

Gene expression studies using microarray enables to monitor transcription of thousands of genes simultaneously. The quantitative level of transcription activity can be assessed under controlled experimental conditions in cell lines or target tissues, and the result can be analyzed using a systematic strategy instead of a gene-by-gene approach. The expression patterns over a large set of genes (genomic profile) can predict different stages of a disease, or discriminate disease cases from normal controls [Gu *et al*., 2002]. Gene expression microarrays analysis can be applied to identify a list of important genes that are critical to the disease manifestation under an environmental condition, and can lead to the identification of etiological pathways.

Compared to the early technology of northern blotting, the microarray technology is improved in two ways: Multiple hybridizations can be carried out simultaneously in a single experiment and therefore experimental variability is reduced; and the transcript abundance can be determined simultaneously for the same RNA samples for all genes, which provides a global view of disease conditions [Duggan *et al*., 1999]

1.2.3.1 Parametric and nonparametric methods for analysis

Parametric modelling attempts to divide the sample variation into unexplained error and known sources. Parameters are estimated to characterize the model and are tested for their significance. These are represented by multivariate regression and ANOVA modelling. They fit the data to a prespecified model with fixed or/and random effects believed to influence expression levels. The difficulty in multivariate modelling of microarrays data lies in the fact that the number of variables (genes) and the number of observations (samples) are unbalanced by magnitudes and we are frequently forced to make simplistic (often untrue) assumptions such as about normality and homoscedasticity among all genes. In addition, current array design often results in the confounding of different types of variations, and makes direct multivariate modelling extremely difficult [Gu *et al*., 2002].

Nonparametric tests are typically more robust than their parametric counterparts as they do not require making strong assumptions on the distribution of the observations. Several nonparametric graphical techniques can be used at the screening stage for data exploration, such as, PCA. The PCA changes the coordinate system using a linear transformation, so that the first few principal components capture most of the variation in the sample. To study the hidden structure inherent in expression patterns, clustering algorithms can be applied to infer ''functional'' grouping of genes and/or samples. The genes can be grouped into similar classes based on their inferred functions, and use the resulting clustering to profile samples for further analysis. Hierarchical clustering (HC) is the most widely applied clustering method in gene expression microarrays data analysis since it assumes no prior knowledge of the data structure. Different schemes can be applied to shape clusters (agglomerative or divisive) and different linkage methods to calculate distances between clusters. In all cases, hierarchical clustering produces a dendogram that reflects somehow the natural ordering of data points [Gu *et al*., 2012].

1.2.3.2 Fold-change

For selecting probes/genes of interest in microarrays analysis is common to calculate fold changes (FC) in normalized expression levels under different conditions, or between cases and control samples. The value of a fold change is calculated as the ratio of expression levels adjusted for background noises on the arrays. A somewhat arbitrary threshold (e.g.,

FC=2, or 3) is then used as a measure to signify the underlying differentiation in RNA expression of the gene. However, the variability in expression levels is correlated with the absolute values of gene expression and therefore this practice is error-prone. Ignoring this nonlinearity and simply using a global FC threshold to distinguish ''false'' from ''real'' differentiation in gene expression could lead to inflated spurious results [Gu *et al*., 2002].

1.2.3.3 Multiple testing and false discovery rate

The problem of multiple testing also exists in gene expression microarray analysis since thousands of genes are tested independently for differential expression and type one error or false positives (i.e., a gene incorrectly identified as differentially expressed) or type II error, or false negative (i.e., failing to identify a truly differentially expressed gene) can become high.

The Bonferroni method, already explained in chapter 1.2.2.3 (page 38), is a simple method to correct for multiple testing that is still widely used in microarray data analysis [Lin, 2005]. This method just divides the P-value cut-off by the number of genes/probes tested. Although this method is quite generally applicable, it is usually not a good choice for microarray studies because it has very low power, i.e., the probability of correctly identifying differentially expressed genes is very small, so many potentially interesting genes may be missed [Verducci *et al*., 2006].

Different criteria to deal with multiple testing have been advocated but the most promising one is the false discovery rate (FDR) [Reiner *et al*., 2003]. FDR is the expected proportion of false positives among all rejected hypotheses. Instead of trying to avoid any false positives, the FDR controls the proportion of positive calls that are false positives. Designing procedures to control the FDR is challenging [Verducci *et al*., 2006]. Three methods frequently used for FDR control are Benjamini and Hochberg (BH), [Benjamini *et al*., 1995] Benjamini-Yekutieli (BY), [Benjamini *et al*., 2001] and the Storey q -value [Storey, 2003; reviewed in Owzar *et al*., 2011].

1.2.3.4 Limitations in gene expression microarrays analysis

Specific problems inherent to the use of microarrays in gene expression studies include: the systematic bias that can be introduced during sample preparation, hybridization and measurement of expression; batch to batch variation in array manufacture, day to day

variation in laboratory conditions and low signal-to-noise ratio in the data [Akey *et al*., 2007; Cookson *et al*., 2009]. Additionally, it was assumed that different microarray platforms give broadly comparable results [Barnes *et al*., 2005]. However, numerous studies are now showing that the overlap in transcript detection between platforms is only ~30–40% [Pedotti *et al*., 2008; van Ruissen *et al*., 2005; Bosotti *et al*., 2007]. The same level of discordance appears when comparisons are made between Affymetrix and Illumina arrays [Barnes *et al*., 2005], Affymetrix and Applied Biosystems arrays [Bosotti *et al*., 2007], or across multiple platforms [Pedotti *et al*., 2008; reviewed in Cookson *et al*., 2009]. Some of this discrepancy may be because individual genes are commonly interrogated by different sequences on different platforms and also derive from the complex and unpredictable factors that determine hybridization of particular nucleic acids to complementary array-bound sequences [Sohail *et al*., 1999; Southern *et al*., 1999] [reviewed in Cookson *et al*., 2009]. A consistent conclusion of comparison studies has been that different platforms provide complementary results [Pedotti *et al.,* 2008; van Ruissen *et al.,* 2005], probably because they are all sampling only a selected fraction of the total transcriptome from the cells or tissue under study [Cookson *et al*., 2009].

A more comprehensive measurement of gene expression comes from arrays that interrogate all known human exons. The use of exon arrays allows the identification of tissue-specific alternative splicing events as well as significant expression outside of known exons and well-annotated genes [Clark *et al*., 2007; Cookson *et al*., 2009].

1.3 REFERENCES

Accardo-Palumbo A, Giardina AR, Ciccia F, Ferrante A, Principato A, Impastato R, Giardina E, Triolo G Phenotype and functional changes of Vgamma9/Vdelta2 T lymphocytes in Behçet's disease and the effect of infliximab on Vgamma9/Vdelta2 T cell expansion, activation and cytotoxicity. Arthritis Res Ther. 2010;12(3):R109.

Adamantiades, B. A case of relapsing iritis with hypopyon (in greek). Archia latrikis Etairias, Athens, pp. 586-593.

Aggarwal S, Ghilardi N, Xie MH, de Sauvage FJ, Gurney AL. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. J Biol Chem. 2003;278(3):1910-4.

Ahmad T, Wallace GR, James T, Neville M, Bunce M, Mulcahy-Hawes K, Armuzzi A, Crawshaw J, Fortune F, Walton R, Stanford MR, Welsh KI, Marshall SE, Jewell DP. Mapping the HLA association in Behçet's disease: a role for tumor necrosis factor polymorphisms? Arthritis Rheum. 2003;48(3):807-13.

Ahmad T, Zhang L, Gogus F, Verity D, Wallace G, Madanat W, Fayyad F, James T, Neville M, Kanawati C, Fortune F, Celik A, Stanford M, Jewell DP, Marshall SE. CARD15 polymorphisms in Behçet's disease. Scand J Rheumatol. 2005;34(3):233-7.

Al-Otaibi LM, Porter SR, Poate TW. Behçet's disease: a review. J Dent Res. 2005; 84(3): 209-22.

Al-Rawi ZS, Sharquie KE, Khalifa SJ, Al-Hadithi FM, Munir JJ. Behçet's disease in Iraqi patients. Ann Rheum Dis. 1986;45(12):987-90.

Amelsfort J, Jacvobs K, Bijlsma J, Lafeber F, Taams L. CD4+ CD25+ regulatory T cells in rheumatoid arthritis. Difference in the presence, phenotype, and function between peripheral blood and synovial fluid. Arthritis Rheum. 2004;50:2775–85.

Akey JM, Biswas S, Leek JT, Storey JD. On the design and analysis of gene expression studies in human populations. Nat Genet. 2007;39(7):807-8; author reply 808-9.

Akman-Demir G, Serdaroglu P, Tasçi B. Clinical patterns of neurological involvement in Behçet's disease: evaluation of 200 patients. The Neuro-Behçet Study Group. Brain. 1999;122(Pt 11): 2171-82.

Aksu K, Kabasakal Y, Sayiner A, Keser G, Oksel F, Bilgiç A, Gümüşdiş G, Doganavşargil E. Prevalences of hepatitis A, B, C and E viruses in Behçet's disease. Rheumatology (Oxford). 1999;38(12): 1279-81.

Arber N, Klein T, Meiner Z, Pras E, Weinberger A. Close association of HLA-B51 and B52 in Israeli patients with Behçet's syndrome. Ann Rheum Dis. 1991;50(6):351-3.

Arida A, Fragiadaki K, Giavri E, Sfikakis PP. Anti-TNF agents for Behçet's disease: analysis of published data on 369 patients. Semin Arthritis Rheum. 2011;41(1):61-70.

Ateş O, Dalyan L, Hatemi G, Hamuryudan V, Topal-Sarikaya A. Genetic susceptibility to Behçet's syndrome is associated with NRAMP1 (SLC11A1) polymorphism in Turkish patients. Rheumatol Int. 2009;29(7):787-91.

Avci O, Ellidokuz E, Simşek I, Büyükgebiz B, Güneş AT. Helicobacter pylori and Behçet's disease. Dermatology. 1999;199(2): 140-3.

Azizleri G, Aksungur VL, Sarica R, Akyol E, Ovül C. The association of HLA-B5 antigen with specific manifestations of Behçet's disease. Dermatology. 1994;188(4):293-5.

Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006;7(10):781-91.

Bang D, Yoon KH, Chung HG, Choi EH, Lee ES, Lee S. Epidemiological and clinical features of Behçet's disease in Korea. Yonsei Med J. 1997;38(6): 428-36.

Bardak Y. Effects of age and sex on Behçet's disease. J Rheumatol. 1999;26(4):1008-9.

Baricordi OR, Sensi A, Pivetti-Pezzi P, Perrone S, Balboni A, Catarinelli G, Filippi F, Melchiorri L, Moncada A, Mattiuz PL. Behcet's disease associated with HLA-B51 and DRw52 antigens in Italians. Hum Immunol. 1986;17(3):297-301.

Barnes M, Freudenberg J, Thompson S, Aronow B, Pavlidis P. Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. Nucleic Acids Res. 2005;33(18):5914-23.

Beck S, Trowsdale J. The human major histocompatability complex: lessons from the DNA sequence. Annu Rev Genomics Hum Genet. 2000;1:117–137

Ben Ahmed M, Houman H, Miled M, Dellagi K, Louzir H. Involvement of chemokines and Th1 cytokines in the pathogenesis of mucocutaneous lesions of Behçet's disease. Arthritis Rheum. 2004;50(7):2291-5.

Behçet H. Uber rezidivierende, apthoese durch ein Virus verursachte Geschwuere am Mund, am Auge und an den Genitalien. Dermatol Wochenschr. 1937;36:1152-1157.

Behçet H, Matteson EL. On relapsing, aphthous ulcers of the mouth, eye and genitalia caused by a virus. 1937. Clin Exp Rheumatol. 2010;28(4 Suppl 60): S2-5.

Benitah NR, Sobrin L, Papaliodis GN. The use of biologic agents in the treatment of ocular manifestations of Behcet's disease. Semin Ophthalmol. 2011;26(4-5):295-303.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B. 1995;17(1):289-300.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statist. 2001; 29(4): 1165–1188.

Benoist C, Mathis D. Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? Nat Immunol. 2001;2(9): 797-801. Review.

Bird Stewart JA. Genetic analysis of families of patients with Behçet's syndrome: data incompatible with autosomal recessive inheritance. Ann Rheum Dis. 1986;45(4):265-8.

Bosotti R, Locatelli G, Healy S, Scacheri E, Sartori L, Mercurio C, Calogero R, Isacchi A. Cross platform microarray analysis for robust identification of differentially expressed genes. BMC Bioinformatics. 2007;8 Suppl 1:S5.

Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet. 2005;366(9489) :941-51.

Chamberlain MA. A family study of Behcet's syndrome. Ann Rheum Dis. 1978;37(5):459-65

Chang JT, Shevach EM, Segal BM. Regulation of interleukin (IL)-12 receptor beta2 subunit expression by endogenous IL-12: a critical step in the differentiation of pathogenic autoreactive T cells. J Exp Med. 1999;189(6):969-78.

Chen Y, Lin CH, Sabatti C. Volume measures for linkage disequilibrium. BMC Genet. 2006;7:54.

Cho SB, Cho S, Bang D. New insights in the clinical understanding of Behçet's disease.Yonsei Med J. 2012;53(1):35-42.

Choukri F, Chakib A, Himmich H, Hüe S, Caillat-Zucman S. HLA-B*51 and B*15 alleles confer predisposition to Behçet's disease in Moroccan patients. Hum Immunol. 2001;62(2):180-5.

Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, Williams A, Blume JE. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 2007;8(4):R64.

Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. Nat Protoc. 2011;6(2):121-33.

Collins A. Approaches to the identification of susceptibility genes. Parasite Immunol. 2009;31(5):225-33.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009;10(3):184-94. Review.

Cordell HJ, Clayton DG. Genetic association studies. Lancet. 2005;366(9491): 1121-31.

Crispin JC, Martinez A, Varela JA. Quantification of regulatory T cells in patients with systemic lupus erythematosus. J Autoimmun. 2003;21:273–6.

Davatchi F, Jamshidi AR, Banihashemi AT, Gholami J, Forouzanfar MH, Akhlaghi M, Barghamdi M, Noorolahzadeh E, Khabazi AR, Salesi M, Salari AH, Karimifar M, Essalat-Manesh K, Hajialiloo M, Soroosh M, Farzad F, Moussavi HR, Samadi F, Ghaznavi K, Asgharifard H, Zangiabadi AH, Shahram F, Nadji A, Akbarian M, Gharibdoost F. WHO-ILAR COPCORD Study (Stage 1, Urban Study) in Iran. J Rheumatol. 2008;35(7): 1384.

Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, Faezi T, Ghodsi Z, Faridar A, Ashofteh F, Sadeghi Abdollahi B. Behcet's disease: from East to West. Clin Rheumatol. 2010a;29(8): 823-33.

Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, Faezi T, Ghodsi Z, Larimi R, Ashofteh F, Abdollahi BS. Behcet's disease in Iran: analysis of 6500 cases. Int J Rheum Dis. 2010b;13(4): 367-73.

Davatchi F. Diagnosis/Classification Criteria for Behcet's Disease. Patholog Res Int. 2012;2012: 607921.

de Menthon M, Lavalley MP, Maldini C, Guillevin L, Mahr A. HLA-B51/B5 and the risk of Behçet's disease: a systematic review and meta-analysis of case-control genetic association studies. Arthritis Rheum. 2009;61(10): 1287-96. Review.

Direskeneli H. Behçet's disease: infectious aetiology, new autoantigens, and HLA-B51. Ann Rheum Dis. 2001;60(11): 996-1002.

Direskeneli H, Saruhan-Direskeneli G. The role of heat shock proteins in Behçet's disease. Clin Exp Rheumatol. 2003;21(4 Suppl 30):S44-8.

Direskeneli H. Autoimmunity vs autoinflammation in Behcet's disease: do we oversimplify a complex disorder? Rheumatology (Oxford). 2006;45(12):1461-5. Review.

Direskeneli H, Fujita H, Akdis CA. Regulation of TH17 and regulatory T cells in patients with Behçet disease. J Allergy Clin Immunol. 2011;128(3):665-6.

Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nat Genet. 1999 Jan;21(1 Suppl):10-4.

Eglin RP, Lehner T, Subak-Sharpe JH. Detection of RNA complementary to herpes-simplex virus in mononuclear cells from patients with Behçet's syndrome and recurrent oral ulcers. Lancet. 1982;2(8312): 1356-61.

Eksioglu-Demiralp E, Direskeneli H, Kibaroglu A, Yavuz S, Ergun T, Akoglu T. Neutrophil activation in Behçet's disease. Clin Exp Rheumatol. 2001;19(5 Suppl 24):S19-24.

Ergun T, Ince U, Ekşioğlu-Demiralp E, Direskeneli H, Gürbüz O, Gürses L, Aker F, Akoğlu T. HSP 60 expression in mucocutaneous lesions of Behçet's disease. J Am Acad Dermatol. 2001;45(6): 904-9.

Frassanito MA, Dammacco R, Cafforio P, Dammacco F. Th1 polarization of the immune response in Behçet's disease: a putative pathogenetic role of interleukin-12. Arthritis Rheum. 1999;42(9): 1967-74.

Fei Y, Webb R, Cobb BL, Direskeneli H, Saruhan-Direskeneli G, Sawalha AH. Identification of novel genetic susceptibility loci for Behçet's disease using a genome-wide association study. Arthritis Res Ther. 2009;11(3): R66.

Feigenbaum A. Description of Behcet's syndrome in the Hippocratic third book of endemic diseases. Br J Ophthalmol. 1956;**40**(6): 355-7.

Freysdottir J, Lau S, Fortune F. Gammadelta T cells in Behçet's disease (BD) and recurrent aphthous stomatitis (RAS). Clin Exp Immunol. 1999;118(3): 451-7.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002;296(5576):2225-9.

Grupo Nacional para o Estudo da Doença de Behçet, Jorge Crespo. Doença de Behçet – Casuística nacional. Medicina interna 1997; 4(4): 225-232.

Geri G, Terrier B, Rosenzwajg M, Wechsler B, Touzot M, Seilhean D, Tran TA, Bodaghi B, Musset L, Soumelis V, Klatzmann D, Cacoub P, Saadoun D. Critical role of IL-21 in modulating TH17 and regulatory T cells in Behçet disease. J Allergy Clin Immunol. 2011;128(3):655-64.

Giordano M. Linkage, Allele sharing and Association. Walker MW, Rapley R, eds. Medical Biomethods Handbook. Humana Press, Totowa, New Jersey 2005; pp: 205-226.

Gu CC, Rao DC, Stormo G, Hicks C, Province MA. Role of gene expression microarray analysis in finding complex disease genes. Genet Epidemiol. 2002;23(1):37-56.

Gül A, Inanç M, Ocal L, Aral O, Koniçe M. Familial aggregation of Behçet's disease in Turkey. Ann Rheum Dis. 2000;59(8): 622-5.

Gül A, Hajeer AH, Worthington J, Barrett JH, Ollier WE, Silman AJ. Evidence for linkage of the HLA-B locus in Behçet's disease, obtained using the transmission disequilibrium test. Arthritis Rheum. 2001;44(1):239-40.

Gül A. Behçet's disease as an autoinflammatory disorder. Curr Drug Targets Inflamm Allergy 2005;4:81–3.

Gunesacar R, Erken E, Bozkurt B, Ozer HT, Dinkci S, Erken EG, Ozbalkan Z. Analysis of CD28 and CTLA-4 gene polymorphisms in Turkish patients with Behcet's disease. Int J Immunogenet. 2007;34(1):45-9.

Hamzaoui K, Hamzaoui A, Hentati F, Kahan A, Ayed K, Chabbou A, Ben Hamida M, Hamza M. Phenotype and functional profile of T cells expressing gamma delta receptor from patients with active Behçet's disease. J Rheumatol. 1994;21(12):2301-6.

Hamzaoui K, Hamzaoui A, Houman H. CD4+CD25+ regulatory T cells in patients with Behçet's disease. Clin Exp Rheumatol. 2006;24(5 Suppl 42):S71-8.

Hamzaoui K, Houman H, Hamzaoui A. Regulatory T cells in cerebrospinal fluid from Behçet's disease with neurological manifestations. J Neuroimmunol. 2007;187(1-2):201-4.

Hamzaoui K, Borhani Haghighi A, Ghorbel IB, Houman H. RORC and Foxp3 axis in cerebrospinal fluid of patients with neuro-Behçet's disease. J Neuroimmunol. 2011;233(1-2):249-53.

Hattersley AT, McCarthy MI. What makes a good genetic association study? Lancet. 2005;366(9493): 1315-23.

Hirohata S. Histopathology of central nervous system lesions in Behçet's disease. J Neurol Sci. 2008;267(1-2):41-7.

Horie Y, Meguro A, Ota M, Kitaichi N, Katsuyama Y, Takemoto Y, Namba K, Yoshida K, Song YW, Park KS, Lee EB, Inoko H, Mizuki N, Ohno S. Association of TLR4 polymorphisms with Behçet's disease in a Korean population. Rheumatology (Oxford). 2009;48(6):638-42.

Horie Y, Meguro A, Kitaichi N, Lee EB, Kanda A, Noda K, Song YW, Park KS, Namba K, Ota M, Inoko H, Mizuki N, Ishida S, Ohno S. Replication of a microsatellite genome-wide association study of Behcet's disease in a Korean population. Rheumatology (Oxford). 2012;51(6):983-6.

Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, Li F, Zhou Q, Ohno S, Chen R, Kijlstra A, Rosenbaum JT, Yang P. Identification of a susceptibility locus in STAT4 for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum. 2012a;64(12): 4104-13.

Hou S, Shu Q, Jiang Z, Chen Y, Li F, Chen F, Kijlstra A, Yang P. Replication study confirms the association between UBAC2 and Behçet's disease in two independent Chinese sets of patients and controls. Arthritis Res Ther. 2012b;29;14(2):R70.

Huang L, Wang C, Rosenberg NA. The relationship between imputation error and statistical power in genetic association studies in diverse populations. Am J Hum Genet. 2009;85(5):692-8.

Hughes T, Coit P, Adler A, Yilmaz V, Aksu K, Düzgün N, Keser G, Cefle A, Yazici A, Ergen A, Alpsoy E, Salvarani C, Casali B, Kötter I, Gutierrez-Achury J, Wijmenga C, Direskeneli H, Saruhan-Direskeneli G, Sawalha AH. Identification of multiple independent susceptibility loci in the HLA region in Behçet's disease. Nat Genet. 2013;45(3):319-24.

Ilter N, Senol E, Gürer MA, Oztaş MO. Behçet's disease and HCV infection. Int J Dermatol. 2000;39(5): 396-7.

International HapMap Consortium. The International HapMap Project. Nature 2003;426: 789-968

International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. Lancet. 1990;335(8697): 1078-80.

International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD). Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 2006;24(suppl 42): S14-S5.

Iwakura Y, Ishigame H. The IL-23/IL-17 axis in inflammation. J Clin Invest. 2006;116(5):1218-22.

Jang WC, Nam YH, Ahn YC, Lee SH, Park SH, Choe JY, Lee SS, Kim SK. Interleukin-17F gene polymorphisms in Korean patients with Behçet's disease. Rheumatol Int. 2008;29(2):173-8.

Jiang Z, Yang P, Hou S, Du L, Xie L, Zhou H, Kijlstra A. L-23R gene confers susceptibility to Behcet's disease in a Chinese Han population. Ann Rheum Dis. 2010;69(7):1325-8.

Juran BD, Lazaridis KN. Applying genomics to the study of complex disease. Semin Liver Dis. 2007 Feb;27(1):3-12.

Kapsimali VD, Kanakis MA, Vaiopoulos GA, Kaklamanis PG. Etiopathogenesis of Behçet's disease with emphasis on the role of immunological aberrations. Clin Rheumatol. 2010;29(11): 1211-6. Review.

Karasneh J, Gül A, Ollier WE, Silman AJ, Worthington J. Whole-genome screening for susceptibility genes in multicase families with Behçet's disease. Arthritis Rheum. 2005a;52(6): 1836-42.

Karasneh JA, Hajeer AH, Silman A, Worthington J, Ollier WE, Gül A. Polymorphisms in the endothelial nitric oxide synthase gene are associated with Behçet's disease. Rheumatology (Oxford). 2005b;44(5):614-7.

Katsantonis J, Adler Y, Orfanos CE, Zouboulis CC. Adamantiades-Behçet's disease: serum IL-8 is a more reliable marker for disease activity than C-reactive protein and erythrocyte sedimentation rate. Dermatology. 2000;201(1):37-9.

Kaya Tİ. Genetics of Behçet's Disease. Patholog Res Int. 2012;2012:912589.

Kera J, Mizuki N, Ota M et al., Significant associations of HLA-B*5101 and B*5108, and lack of association of class II alleles with Behcet's disease in Italian patients. Tissue Antigens 1999;54:565–71

Kilmartin DJ, Finch A, Acheson RW. Primary association of HLA-B51 with Behçet's disease in Ireland. Br J Ophthalmol. 1997;81(8):649-53.

Kim MS, Kim JH. Prognostic comparison of Behçet's disease with or without HLA-Bw 51 antigen. Korean J Ophthalmol. 1989;3(2):85-9.

Kim EH, Mok JW, Bang DS, Lee ES, Lee SN, Park KS. Intercellular adhesion molecule-1 polymorphisms in Korean patients with Behcet's disease. J Korean Med Sci. 2003;18(3):415-8.

Kim SK, Jang WC, Park SB, Park DY, Bang KT, Lee SS, Jun JB, Yoo DH, Chang HK. SLC11A1 gene polymorphisms in Korean patients with Behçet's disease. Scand J Rheumatol. 2006;35(5):398-401.

Kim J, Park JA, Lee EY, Lee YJ, Song YW, Lee EB. Imbalance of Th17 to Th1 cells in Behçet's disease. Clin Exp Rheumatol. 2010;28(4 Suppl 60):S16-9.

Kim JR, Chae JN, Kim SH, Ha JS. Subpopulations of regulatory T cells in rheumatoid arthritis, systemic lupus erythematosus, and Behcet's disease. J Korean Med Sci. 2012;27(9):1009-13.

Kiraz S, Oztürk MA, Ertenli I, Calgüneri M. Parvovirus B19 infection in Behçet's disease. Ann Rheum Dis. 2001;60(8): 814-5.

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sacli FS, Erer B, Inoko H, Emrence Z, Cakar A, Abaci N, Ustek D, Satorius C, Ueda A, Takeno M, Kim Y, Wood GM, Ombrello MJ, Meguro A, Gül A, Remmers EF, Kastner DL. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013;45(2):202-7.

Koç Y, Güllü I, Akpek G, Akpolat T, Kansu E, Kiraz S, Batman F, Kansu T, Balkanci F, Akkaya S, et al., Vascular involvement in Behçet's disease. J Rheumatol. 1992;19(3):402-10.

Koné-Paut I, Yurdakul S, Bahabri SA, Shafae N, Ozen S, Ozdogan H, Bernard JL. Clinical features of Behçet's disease in children: an international collaborative study of 86 cases. J Pediatr. 1998;132(4): 721-5.

Koné-Paut I, Geisler I, Wechsler B, Ozen S, Ozdogan H, Rozenbaum M, Touitou I. Familial aggregation in Behçet's disease: high frequency in siblings and parents of pediatric probands. J Pediatr. 1999;135(1): 89-93.

Koné-Paut I, Gorchakoff-Molinas A, Weschler B, Touitou I. Paediatric Behçet's disease in France. Ann Rheum Dis. 2002;61(7): 655-6.

Köse O. Development of Immunopathogenesis Strategies to Treat Behçet's Disease. Patholog Res Int. 2012;2012:261989.

Kötter I, Günaydin I, Stübiger N, Yazici H, Fresko I, Zouboulis CC, Adler Y, Steiert I, Kurz B, Wernet D, Braun B, Müller CA. Comparative analysis of the association of HLA-B*51 suballeles with Behçet's disease in patients of German and Turkish origin. Tissue Antigens. 2001;58(3):166-70.

Krause I, Uziel Y, Guedj D, Mukamel M, Harel L, Molad Y, Weinberger A. Childhood Behçet's disease: clinical features and comparison with adult-onset disease. Rheumatology (Oxford). 1999;38(5): 457-62.

Kural-Seyahi E, Fresko I, Seyahi N, Ozyazgan Y, Mat C, Hamuryudan V, Yurdakul S, Yazici H. The long-term mortality and morbidity of Behçet syndrome: a 2-decade outcome survey of 387 patients followed at a dedicated center. Medicine (Baltimore). 2003;82(1): 60-76.

Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, McClanahan T, Kastelein RA, Cua DJ. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. J Exp Med. 2005;201(2):233-40.

Lamb JR, Young DB. T cell recognition of stress proteins. A link between infectious and autoimmune disease. Mol Biol Med. 1990;7(4):311-21.

Lehner T, Batchelor JR, Challacombe SJ, Kennedy L. An immunogenetic basis for the tissue involvement in Behçet's syndrome. Immunology. 1979;37(4):895-900.

Lehner T, Lavery E, Smith R, van der Zee R, Mizushima Y, Shinnick T. Association between the 65-kilodalton heat shock protein, Streptococcus sanguis, and the corresponding antibodies in Behçet's syndrome. Infect Immun. 1991;59(4): 1434-41.

Lehner T. The role of heat shock protein, microbial and autoimmune agents in the aetiology of Behçet's disease. Int Rev Immunol. 1997;14(1):21-32.

Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, Song R, Kang YM, Kang SW, Baek HJ, Kitaichi N, Meguro A, Mizuki N, Namba K, Ishida S, Kim J, Niemczek E, Lee EY, Song YW, Ohno S, Lee EB. Genome-wide association study identifies GIMAP as a novel susceptibility locus for Behcet's disease. Ann Rheum Dis. 2013 Feb 25. [Epub ahead of print]

Li K, Zhao M, Hou S, Du L, Kijlstra A, Yang P. Association between polymorphisms of FCRL3, a non-HLA gene, and Behçet's disease in a Chinese population with ophthalmic manifestations. Mol Vis. 2008;14:2136-42.

Liang SC, Tan XY, Luxenberg DP, Karim R, Dunussi-Joannopoulos K, Collins M, Fouser LA. Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. J Exp Med. 2006;203(10):2271-9.

Lin DY. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics. 2005;21(6):781-7.

Maldini C, Lavalley MP, Cheminant M, de Menthon M, Mahr A. Relationships of HLA-B51 or B5 genotype with Behcet's disease clinical characteristics: systematic review and meta-analyses of observational studies. Rheumatology (Oxford). 2012;51(5): 887-900. Review.

Manetti R, Parronchi P, Giudizi MG, Piccinni MP, Maggi E, Trinchieri G, Romagnani S. Natural killer cell stimulatory factor (interleukin 12 [IL-12]) induces T helper type 1 (Th1)-specific immune responses and inhibits the development of IL-4-producing Th cells. J Exp Med. 1993;177(4):1199-204.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363(2):166-76.

Marshall SE. Behçet's disease. Best Pract Res Clin Rheumatol. 2004; 18(3): 291-311. Review.

Matsui T, Kurokawa M, Kobata T, Oki S, Azuma M, Tohma S, Inoue T, Yamamoto K, Nishioka K, Kato T. Autoantibodies to T cell costimulatory molecules in systemic autoimmune diseases. J Immunol. 1999;162(7):4328-35.

Meguro A, Ota M, Katsuyama Y, Oka A, Ohno S, Inoko H, Mizuki N. Association of the toll-like receptor 4 gene polymorphisms with Behcet's disease. Ann Rheum Dis. 2008;67(5):725-7.

Meguro A, Inoko H, Ota M, Katsuyama Y, Oka A, Okada E, Yamakawa R, Yuasa T, Fujioka T, Ohno S, Bahram S, Mizuki N. Genetics of Behçet disease inside and outside the MHC. Ann Rheum Dis. 2010;69(4): 747-54.

Mendoza-Pinto C, García-Carrasco M, Jiménez-Hernández M, Jiménez Hernández C, Riebeling-Navarro C, Nava Zavala A, Vera Recabarren M, Espinosa G, Jara Quezada J, Cervera R. Etiopathogenesis of Behcet's disease. Autoimmun Rev. 2010;9(4): 241-5.

Mizuki N, Meguro A, Tohnai I, Gül A, Ohno S, Mizuki N. Association of Major Histocompatibility Complex Class I Chain-Related Gene A and HLA-B Alleles with Behçet's Disease in Turkey. Jpn J Ophthalmol. 2007;51(6):431-6.

Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, Ito N, Kera J, Okada E, Yatsu K, Song YW, Lee EB, Kitaichi N, Namba K, Horie Y, Takeno M, Sugita S, Mochizuki M, Bahram S, Ishigatsubo Y, Inoko H. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat Genet. 2010;42(8): 703-6.

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol. 2006;241(2):252-61.

Muhaya M, Lightman S, Ikeda E, Mochizuki M, Shaer B, McCluskey P, Towler HM. Behçet's disease in Japan and in Great Britain: a comparative study. Ocul Immunol Inflamm. 2000;8(3): 141-8.

Mumcu G, Inanc N, Yavuz S, Direskeneli H. The role of infectious agents in the pathogenesis, clinical manifestations and treatment strategies in Behçet's disease. Clin Exp Rheumatol. 2007;25(4 Suppl 45): S27-33.

Nam EJ, Han SW, Kim SU, Cho JH, Sa KH, Lee WK, Park JY, Kang YM. Association of vascular endothelial growth factor gene polymorphisms with behcet disease in a Korean population. Hum Immunol. 2005;66(10):1068-73.

Nanke Y, Kotake S, Goto M, Ujihara H, Matsubara M, Kamatani N. Decreased percentages of regulatory T cells in peripheral blood of patients with Behcet's disease before ocular attack: a possible predictive marker of ocular attack. Mod Rheumatol. 2008;18(4):354-8.

Nomura Y, Kitteringham N, Shiba K, Goseki M, Kimura A, Mineshita S. Use of the highly sensitive PCR method to detect the Herpes simplex virus type 1 genome and its expression in samples from Behçet disease patients. J Med Dent Sci. 1998;45(1): 51-8.

Ohno S, Ohguchi M, Hirose S, Matsuda H, Wakisaka A, Aizawa M. Close association of HLA-Bw51 with Behçet's disease. Arch Ophthalmol. 1982;100(9): 1455-8.

Oksel F, Keser G, Ozmen M, Aksu K, Kitapcioglu G, Berdeli A, Doganavsargil E. Endothelial nitric oxide synthase gene Glu298Asp polymorphism is associated with Behçet's disease. Clin Exp Rheumatol. 2006;24(5 Suppl 42):S79-82.

Owzar K, Barry WT, Jung SH. Statistical considerations for analysis of microarray experiments. Clin Transl Sci. 2011 Dec;4(6):466-77.

Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet. 2005;366(9492): 1223-34.

Pay S, Simşek I, Erdem H, Dinç A. Immunopathogenesis of Behçet's disease with special emphasize on the possible role of antigen presenting cells. Rheumatol Int. 2007;27(5):417-24. Review.

Parham C, Chirica M, Timans J, Vaisberg E, Travis M, Cheung J, Pflanz S, Zhang R, Singh KP, Vega F, To W, Wagner J, O'Farrell AM, McClanahan T, Zurawski S, Hannum C, Gorman D, Rennick DM, Kastelein RA, de Waal Malefyt R, Moore KW. A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R. J Immunol. 2002;168(11):5699-708.

Park K, Kim N, Nam J, Bang D, Lee ES. Association of TNFA promoter region haplotype in Behçet's Disease. J Korean Med Sci. 2006;21(4):596-601.

Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA. 2008;299(11):1335-44.

Pedotti P, 't Hoen PA, Vreugdenhil E, Schenk GJ, Vossen RH, Ariyurek Y, de Hollander M, Kuiper R, van Ommen GJ, den Dunnen JT, Boer JM, de Menezes RX. Can subtle changes in gene expression be consistently detected with different microarray platforms? BMC Genomics. 2008;9:124.

Piga M, Mathieu A. Genetic susceptibility to Behcet's disease: role of genes belonging to the MHC region. Rheumatology (Oxford). 2011;50(2):299-310. Review.

Pineton de Chambrun M, Wechsler B, Geri G, Cacoub P, Saadoun D. New insights into the pathogenesis of Behçet's disease. Autoimmun Rev. 2012;11(10):687-98.

Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet. 2001;69(1):1-14.

Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics. 2003;19(3):368-75.

Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, Le JM, Yang B, Korman BD, Cakiris A, Aglar O, Emrence Z, Azakli H, Ustek D, Tugal-Tutkun I, Akman-Demir G, Chen W, Amos CI, Dizon MB, Kose AA, Azizlerli G, Erer B, Brand OJ, Kaklamani VG, Kaklamanis P, Ben-Chetrit E, Stanford M, Fortune F, Ghabra M, Ollier WE, Cho YH, Bang D, O'Shea J, Wallace GR, Gadina M, Kastner DL, Gül A. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet. 2010;42(8): 698-702.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001;69(1):138-47.

Roberts R, Wells GA, Stewart AF, Dandona S, Chen L. The genome-wide association study--a new era for common polygenic disorders. J Cardiovasc Transl Res. 2010;3(3):173-82.

Russell AI, Lawson WA, Haskard DO. Potential new therapeutic options in Behçet's syndrome. BioDrugs. 2001;15(1): 25-35.

Salvarani C, Boiardi L, Casali B, Olivieri I, Ciancio G, Cantini F, Salvi F, Malatesta R, Govoni M, Trotta F, Filippini D, Paolazzi G, Nicoli D, Farnetti E, Macchioni P. Endothelial nitric oxide synthase gene polymorphisms in Behçet's disease. J Rheumatol. 2002;29(3):535-40.

Sakane T, Suzuki N, Nagafuchi H. Etiopathology of Behçet's disease: immunological aspects. Yonsei Med J. 1997;38(6): 350-8. Review.

Sakane T, Takeno M, Suzuki N, Inaba G. Behçet's disease. N Engl J Med. 1999;341(17): 1284-91. Review.

Saylan T, Mat C, Fresko I, Melikoğlu M. Behçet's disease in the Middle East. Clin Dermatol. 1999;17(2): 209-23; discussion 105-6. Review.

Sohail M, Akhtar S, Southern EM. The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides. RNA. 1999;5(5):646-55.

Song YW, Kang EH. Behçet's disease and genes within the major histocompatibility complex region. Mod Rheumatol. 2012;22(2):178-85. Review.

Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. Nat Genet. 1999;21(1 Suppl):5-9.

Soylu M, Ersöz TR, Erken E. The association between HLA B5 and ocular involvement in Behçet's disease in southern Turkey. Acta Ophthalmol (Copenh). 1992;70(6):786-9.

Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Statist. 2003; 31(6):2013–2035.

Stram DO and Seshan VE. Multi-SNP Haplotype Analysis Methods for Association Analysis. Methods Mol Biol. 2012; 850:423-52.

Studd M, McCance DJ, Lehner T. Detection of HSV-1 DNA in patients with Behçet's syndrome and in patients with recurrent oral ulcers by the polymerase chain reaction. J Med Microbiol. 1991;34(1): 39-43.

Takeno M, Kariyone A, Yamashita N, Takiguchi M, Mizushima Y, Kaneoka H, Sakane T. Excessive function of peripheral blood neutrophils from patients with Behçet's disease and from HLA-B51 transgenic mice. Arthritis Rheum. 1995;38(3): 426-33.

Tunes R, Santiago M. Behcet's Syndrome: Literature Review. Current Rheumatology Reviews 2009;5: 64-82.

Turan B, Gallati H, Erdi H, Gürler A, Michel BA, Villiger PM. Systemic levels of the T cell regulatory cytokines IL-10 and IL-12 in Bechçet's disease; soluble TNFR-75 as a biological marker of disease activity. J Rheumatol. 1997;24(1):128-32.

Tursen U, Gurler A, Boyvat A. Evaluation of clinical findings according to sex in 2313 Turkish patients with Behçet's disease. Int J Dermatol. 2003;42(5): 346-51.

Uzunoğlu S, Acar H, Okudan N, Gökbel H, Mevlitoğlu I, Sari F. Evaluation of the association between null genotypes of glutathione-S-transferases and Behcet's disease. Arch Dermatol Res. 2006;297(7):289-93.

van Ruissen F, Ruijter JM, Schaaf GJ, Asgharnegad L, Zwijnenburg DA, Kool M, Baas F. Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. BMC Genomics. 2005;6:91.

Vens M, Ziegler A. Estimating disequilibrium coefficients. Methods Mol Biol. 2012;850:103-17.

Verducci JS, Melfi VF, Lin S, Wang Z, Roy S, Sen CK. Microarray analysis of gene expression: considerations in data mining and statistical treatment. Physiol Genomics. 2006;25(3):355-63.

Verity DH, Marr JE, Ohno S, Wallace GR, Stanford MR. Behçet's disease, the Silk Road and HLA-B51: historical and geographical perspectives. Tissue Antigens. 1999a;54(3):213-20.

Verity DH, Wallace GR, Vaughan RW, Kondeatis E, Madanat W, Zureikat H, Fayyad F, Marr JE, Kanawati CA, Stanford MR. HLA and tumour necrosis factor (TNF) polymorphisms in ocular Behçet's disease. Tissue Antigens. 1999b;54(3):264-72.

Verity DH, Vaughan RW, Kondeatis E, Madanat W, Zureikat H, Fayyad F, Marr JE, Kanawati CA, Wallace GR, Stanford MR. Intercellular adhesion molecule-1 gene polymorphisms in Behçet's disease. Eur J Immunogenet. 2000;27(2):73-6.

Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7-24.

Wallace GR, Kondeatis E, Vaughan RW, Verity DH, Chen Y, Fortune F, Madanat W, Kanawati CA, Graham EM, Stanford MR. IL-10 genotype analysis in patients with Behçet's disease. Hum Immunol. 2007;68(2):122-7.

Whitcomb DC, Aoun E, Vodovotz Y, Clermont G, Barmada MM. Evaluating disorders with a complex genetics basis. the future roles of meta-analysis and systems biology. Dig Dis Sci. 2005;50(12):2195-202.

Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, Abdollahi BS, Nadji A, Jesus G, Barcelos F, Patto JV, Shafiee NM, Ghaderibarim F, Oliveira SA. Association study of IL10 and IL23R-IL12RB2 in Iranian patients with Behçet's disease. Arthritis Rheum. 2012;64(8):2761-72.

Yalçindağ, Uzun A. Results of interferon alpha-2a therapy in patients with Behcet's disease. J Ocul Pharmacol Ther. 2012;28(4):439-43.

Yazici H, Tüzün Y, Pazarli H, Yurdakul S, Ozyazgan Y, Ozdoğan H, Serdaroğlu S, Ersanli M, Ulkü BY, Müftüoğlu AU. Influence of age of onset and patient's sex on the prevalence and severity of manifestations of Behçet's syndrome. Ann Rheum Dis. 1984;43(6):783-9.

Yazici H, Başaran G, Hamuryudan V, Hizli N, Yurdakul S, Mat C, Tüzün Y, Ozyazgan Y, Dimitriyadis I. The ten-year mortality in Behçet's syndrome. Br J Rheumatol. 1996;35(2):139-41.

Yazici H. The place of Behçet's syndrome among the autoimmune diseases. Int Rev Immunol. 1997;14(1): 1-10.

Yazici H, Ozyazgan Y. Medical management of Behçet's syndrome. Dev Ophthalmol. 1999;31: 118-31. Review.

Yazici H. The lumps and bumps of Behçet's syndrome. Autoimmun Rev. 2004;3 Suppl 1: S53-4.

Zouboulis CC, Kötter I, Djawari D, Kirch W, Kohl PK, Ochsendorf FR, Keitel W, Stadler R, Wollina U, Proksch E, Söhnchen R, Weber H, Gollnick HP, Hölzle E, Fritz K, Licht T, Orfanos CE. Epidemiological features of Adamantiades-Behçet's disease in Germany and in Europe. Yonsei Med J. 1997;38(6): 411-22.

Zouboulis CC, Katsantonis J, Ketteler R, Treudler R, Kaklamani E, Hornemann S, Kaklamanis P, Orfanos CE. Adamantiades-Behçet's disease: interleukin-8 is increased in serum of patients with active oral and neurological manifestations and is secreted by small vessel endothelial cells. Arch Dermatol Res. 2000;292(6): 279-84.

Zouboulis CC. Benediktos Adamantiades and his forgotten contributions to medicine. Eur J Dermatol. 2002;12(5): 471-4.

Zouboulis CC, May T. Pathogenesis of Adamantiades-Behçet's disease. Med Microbiol Immunol. 2003;192(3): 149-55.

CHAPTER 2

Aim of the thesis

The aim of this research project is to unravel novel genetic risk factors implicated in Behçet's disease susceptibility and to fine-map recent genes found associated with BD, using a large and well characterized dataset of Iranian BD cases and controls.

Therefore, the specific aims of this thesis are:

1) To assess the existence of population stratification in the Iranian dataset, and control for if necessary, using a panel of ancestry informative markers;
2) To investigate the association of the mitochondrial haplogroups and variants with BD in the Iranian population;
3) To replicate and fine-map the association of genes recently pointed out as BD genetic risk factors, in genome-wide association studies, performed in the Turkish and Japanese population, such as, the *IL10* and *IL12RB2-IL23R* genes;
4) To perform a gene expression study in Portuguese BD cases and controls and to test the association of the differential expressed genes, with BD, in the Iranian population;
5) To perform the first whole genome association scan in the Iranian population using a DNA pooling approach;
6) To confirm our positive findings in independent datasets.

CHAPTER 3

Association of mitochondrial polymorphism m.709G>A with Behçet's disease

(Xavier *et al.*, Ann Rheum Dis. 2011)

3.1 MAIN MANUSCRIPT

Association of mitochondrial polymorphism m.709G>A with Behçet's disease

The involvement of nuclear genes in Behçet's disease (BD) risk has been investigated, but the role of the mitochondrial DNA (mtDNA) has been completely neglected. Mitochondria are the main intracellular source of reactive oxygen species produced during normal aerobic metabolism via the electron transport chain and since mitochondrial dysfunction may underlie a multitude of clinical features in multifactorial and multisystemic diseases such as BD, we assessed whether mtDNA single nucleotide polymorphisms (SNPs) and haplogroups confer susceptibility to BD.

A total of 615 patients and 434 controls from Iran were enrolled in this study. BD diagnosis was made according to the revised International Criteria for Behçet's Disease[1] (ICBD cases). A total of 494 patients also fulfilled the International Study Group[2] criteria for diagnosis of BD (ISG cases). We genotyped 19 mtDNA SNPs which are sufficient for classifying our Iranian cohort into their most prevalent haplogroups: West Eurasian R0, H, V, J, T, U, K, N1, N1e'I, I, X and W haplogroups; Eastern Eurasian macrohaplogroup M, D, N (except for haplogroups N1, N1e'I, I, W and X) and R (lineages except R0, JT and UK) haplogroups; and African L haplogroup (figure 1).[3–6] Using a panel of 89 autosomal ancestry informative markers, no evidence of population admixture was found in our cohort

	SNP ID	m.709G>A	m.1719G>A	m.4216T>C	m.4580G>A	m.5178C>A	m.7028T>C	m.8251G>A	m.8701A>G	m.9055G>A	m.10034T>C	m.10398A>G	m.10400C>T	m.10873T>C	m.11719G>A	m.12308A>G	m.12705C>T	m.13368G>A	m.14470T>C	m.15607A>G	
	SNP type	ncod	ncod	p.Y304H	syn	p.L237M	syn	syn	p.T59A	p.A177T	ncod	p.T114A	syn	syn	syn	ncod	syn	syn	syn	syn	Haplogroup frequency (%)
Haplogroup	H						**C**								**G**		**C**				16.4
	R0				**G**		**T**								**G**		C				13.6
	V				**A**		T								**G**						0.0
	J	**G**	G	**C**								**G**			A		**C**	**G**			14.0
	T	**A**		**C**			T					**A**			A		C	**A**		**G**	9.1
	U						T		**A**	**G**				**T**		**G**	C			A	20.6
	K								**A**	**A**						**G**	**C**			A	6.5
	R*					**C**			**A**					**T**	**A**	**A**	**C**			A	2.1
	N*		**G**					**G**	**A**		**T**					**A**	**T**		**T**		1.2
	N1		**A**					**G**	**A**		**T**	**A**					T				0.9
	N1e'I		**A**						**A**		**T**						T				2.1
	I		**A**					**A**	**A**		**C**	**G**					T				3.0
	W	**A**	**G**					**A**	**A**			**A**				**A**	T				2.6
	X								**A**								T		**C**		1.6
	L								**G**		**T**	**G**	**C**	**C**			T				0.7
	M*					**C**			**G**		**T**		**T**	**C**			T				4.0
	D					**A**							**T**	**C**							1.6
SNP frequency (%)	Controls	14.5	7.0	23.4	0.0	1.6	83.2	7.5	6.3	7.0	3.0	29.0	5.6	6.3	70.1	27.1	17.5	9.3	1.9	9.3	
	ICBD cases	19.6	7.2	24.5	0.3	1.5	80.9	7.9	5.1	7.9	2.6	28.0	4.6	5.2	69.7	25.5	16.5	11.3	2.1	11.0	
	ISG cases	21.4	6.9	25.5	0.4	1.2	82.3	8.4	4.9	7.3	2.2	27.3	4.3	4.9	70.5	24.6	17.1	12.0	2.4	11.6	
ICBD cases vs controls	p (unadjusted)	**0.038**	0.980	0.714	0.641	0.963	0.299	0.914	0.474	0.696	0.832	0.782	0.548	0.550	0.952	0.631	0.778	0.365	0.945	0.458	
	OR (95% CI)	1.46 (1.03 to 2.02)																			
	p (adjusted)	0.053	0.997	0.665	0.972	0.859	0.296	0.667	0.436	0.635	0.843	0.836	0.514	0.517	0.959	0.562	0.862	0.394	0.856	0.503	
	OR (95% CI)	1.40 (1.00 to 1.97)																			
ISG cases vs controls	p (unadjusted)	**0.007**	0.924	0.459	0.537	0.807	0.690	0.702	0.442	0.905	0.595	0.683	0.439	0.446	0.784	0.449	0.972	0.213	0.704	0.292	
	OR (95% CI)	1.61 (1.14 to 2.27)																			
	p (adjusted)	**0.013**	0.890	0.505	0.971	0.759	0.585	0.536	0.395	0.839	0.579	0.671	0.413	0.437	0.867	0.397	0.981	0.238	0.585	0.330	
	OR (95% CI)	1.56 (1.10 to 2.21)																			

R*: all lineages inside haplogroup R except for haplogroups JT and UK; N*: all lineages inside haplogroup N except for haplogroups N1, I, W and X; M*: macro-haplogroup M except for haplogroup D.

Figure 1 Characterisation and association of the investigated mitochondrial markers and haplogroups. Each haplogroup was determined by the combination of alleles in bold, and the alleles not in bold aided in the phylogenetic assignment. The polymorphisms are named after their base pair position and alleles. The type of the variant is indicated as 'ncod' for non-coding single nucleotide polymorphisms (SNPs), 'syn' for synonymous SNPs, and the amino acid substitution is also shown for non-synonymous SNPs. The haplogroup frequencies in controls and the SNP frequencies of the derived allele (second allele in the SNP ID) are indicated. The results of mitochondrial SNP association testing with Behçet's disease risk using the ICBD cases or ISG cases are shown. Unadjusted and gender-adjusted p values are presented and significant p values (<0.05) are highlighted in bold. ORs and 95% CIs are shown only for allele A of m.709G>A. ICBD, International Criteria for Behçet's Disease; ISG, International Study Group Behçet's Disease.

(data not shown). Extensive genotyping quality control checks were implemented.

The associations with BD risk were assessed using Pearson's χ^2 tests and logistic regression analyses with gender as a covariate. Each haplogroup was compared with all other haplogroups pooled together. Results were considered significant below the conventional level of p=0.05. Since some of the markers are in linkage disequilibrium and the haplogroup comparisons are not independent, we did not perform corrections for multiple testing and uncorrected p values are reported.

The general characteristics (table 1) and the observed haplogroup frequencies (figure 1) in our cohort are in agreement with those previously reported in the Iranian population.[7 8] In the ICBD dataset, m.709G>A (A allele in 14.5% of controls, 19.6% of ICBD cases) was significantly associated with BD prior (p=0.038) and marginally after adjustment for gender (p=0.053). In ISG cases, m.709G>A was also the only marker associated with BD prior (p=0.007, A allele in 21.4% of ISG cases) and after adjustment for gender (p=0.013). This marker does not define any haplogroup by itself. None of the haplogroups tested showed an association with BD risk.

Our findings link the mtDNA m.709G>A non-coding variant in the 12S rRNA (MT-RNR1) locus with risk for BD. Consistent with the multisystemic nature and the gene–environmental interaction of BD, ribosomal RNAs are among the very few genes present in all cells. 12S rRNA is a 959-nucleotide molecule participating in the assembly of amino acids into functional proteins. Mutations in MT-RNR1 are known to cause maternally inherited non-syndromic and antibiotic-induced deafness. Further investigation is required to determine if m.709G>A contributes to BD susceptibility by impairing the ability of mitochondria to produce proteins and enhancing oxidative stress or through another mechanism. Increased oxidative stress and impaired antioxidant defence system observed in blood of patients with BD[9 10] may significantly contribute to the disease pathophysiology.

Interestingly, we found a stronger association of mtDNA SNP m.709G>A with BD in the subset of samples fulfilling the more specific ISG criteria than in the entire dataset of BD cases, supporting the notion that the ISG subset may be a genetically more homogeneous group of cases.

Table 1 General characteristics of the study sample

Characteristic	Controls	ICBD cases	ISG cases
N	434	615	494
Gender, n/N (% males)	407/434 (93.8)	498/615 (81.0)	409/494 (82.8)
Mean age at examination, years± SD	41.4±11.6	38.7±10.4	39.2±10.7
Mean age at diagnosis, years± SD	–	31.5±8.5	31.9±8.9
Oral aphtosis, n/N (%)	92/434 (21.2)	607/615 (98.7)	494/494 (100)
Genital aphtosis, n/N (%)	0/434 (0.0)	372/615 (60.5)	343/494 (69.4)
Skin lesions, n/N (%)		364/615 (59.2)	357/494 (72.3)
Pseudofolliculitis		290/615 (47.2)	286/494 (57.9)
Erythema nodosum		133/615 (21.6)	129/494 (26.1)
Skin aphtosis		18/615 (2.9)	17/494 (3.4)
Ophthalmological manifestations, n/N (%)		399/615 (64.9)	312/494 (63.2)
Anterior uveitis		290/615 (47.2)	221/494 (44.7)
Posterior uveitis		341/615 (55.4)	264/494 (53.4)
Retinal vasculitis		235/615 (38.2)	192/494 (38.9)
Joint manifestations, n/N (%)		205/615 (33.3)	175/494 (35.4)
Arthralgia		84/615 (13.7)	66/494 (13.4)
Arthritis		133/615 (21.6)	120/494 (24.3)
Ankylosing spondylitis		12/615 (2.0)	8/494 (1.6)
Neurological manifestations, n/N (%)		38/615 (6.2)	33/494 (6.7)
Vascular involvement, n/N (%)		38/615 (6.2)	33/494 (6.7)
Gastrointestinal manifestations, n/N (%)		17/615 (2.8)	16/494 (3.2)
Epididymitis, n/N (%)		21/498 (4.2)	18/409 (4.4)
Cardiac involvement, n/N (%)		4/615 (0.7)	4/494 (0.8)
Pleuropulmonary involvement, n/N (%)		5/615 (0.8)	5/494 (1.0)
Pathergy phenomenon, n/N (%)		281/603 (46.6)	275/483 (56.9)
Family history of Behçet's disease, n/N (%)		52/586 (8.9)	40/468 (8.5)

ICBD, International Criteria for Behçet's Disease; ISG, International Study Group Behçet's Disease.

This study links for the first time the mtDNA with BD susceptibility and suggests that, in addition to multiple nuclear genes and environmental contributions, BD risk might also be governed by mitochondrial genomic background. This preliminary association warrants further validation and investigation in BD and in other rheumatic diseases.

Joana M Xavier,[1,2] Niloofar Mojarad Shafiee,[3] Fahmida Ghaderi,[3] Alexandra Rosa,[4] Bahar Sadeghi Abdollahi,[3] Abdolhadi Nadji,[3] Farhad Shahram,[3] Fereydoun Davatchi,[3] Sofia A Oliveira[1,2]

[1]Instituto de Medicina Molecular, Lisboa, Portugal
[2]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[3]Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran
[4]Unidade Ciências Médicas, Centro Competências das Ciências da Vida, Universidade da Madeira, Funchal, Portugal

Correspondence to Sofia A Oliveira, Instituto de Medicina Molecular, Avenida Professor Egas Moniz, Edificio Egas Moniz, 1649-028 Lisboa, Portugal; aaoliveira@fm.ul.pt

Acknowledgements We thank Dr Majid Zeidi, Iranian Blood Transfusion Organization, for his valuable support. We also thank to Doctor Sirous Zeinali, Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran and Doctor Kayvan Saeedfar for their valuable help. We are also deeply grateful to all study participants and to the genotyping unit at the Instituto Gulbenkian de Ciência.

Funding This work was supported by the grant from the Research Committee of the Tehran University of Medical Sciences under the registration number 132/714, the PTDC/SAU-GMG/098937/2008 grant and a doctoral fellowship (JMX) from the Portuguese Fundação para a Ciência e a Tecnologia, and a fellowship from the Portuguese Instituto do Emprego e Formação Profissional (JMX).

Competing interests None.

Patient consent Obtained.

Ethics approval This study received ethics approval from the ethics committees at the Rheumatology Center, Tehran University for Medical Sciences, Iran, and from the Portuguese Institute of Rheumatology, Lisbon, Portugal.

Provenance and peer review Not commissioned; externally peer reviewed.

Accepted 23 January 2011
Published Online First 22 February 2011

Ann Rheum Dis 2011;**70**:1514–1516.
doi:10.1136/ard.2010.143537

REFERENCES

1. **International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD)**. Revision of the International Criteria for Behçet's Disease (ICBD). *Clin Exp Rheumatol* 2006;**24**(Suppl 42):S14–15.
2. **International Study Group for Behçet's Disease**. Criteria for diagnosis of Behçet's disease. *Lancet* 1990;**335**:1078–80.
3. **Torroni A**, Huoponen K, Francalacci P, *et al*. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996;**144**:1835–50.
4. **Torroni A**, Bandelt HJ, Macaulay V, *et al*. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 2001;**69**:844–52.
5. **Richards M**, Macaulay V, Hickey E, *et al*. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000;**67**:1251–76.
6. **Macaulay V**, Richards M, Hickey E, *et al*. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999;**64**:232–49.
7. **Davatchi F**, Shahram F, Chams-Davatchi C, *et al*. Behcet's disease: from East to West. *Clin Rheumatol* 2010;**29**:823–33.
8. **Al-Zabery N**, Semino O, Benuzzi G, *et al*. Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 2003;**28**:458–72.
9. **Taysi S**, Demircan B, Akdeniz N, *et al*. Oxidant/antioxidant status in men with Behçet's disease. *Clin Rheumatol* 2007;**26**:418–22.
10. **Freitas JP**, Filipe P, Yousefi A, *et al*. Oxidative stress in Adamantiades-Behçet's disease. *Dermatology (Basel)* 1998;**197**:343–8.

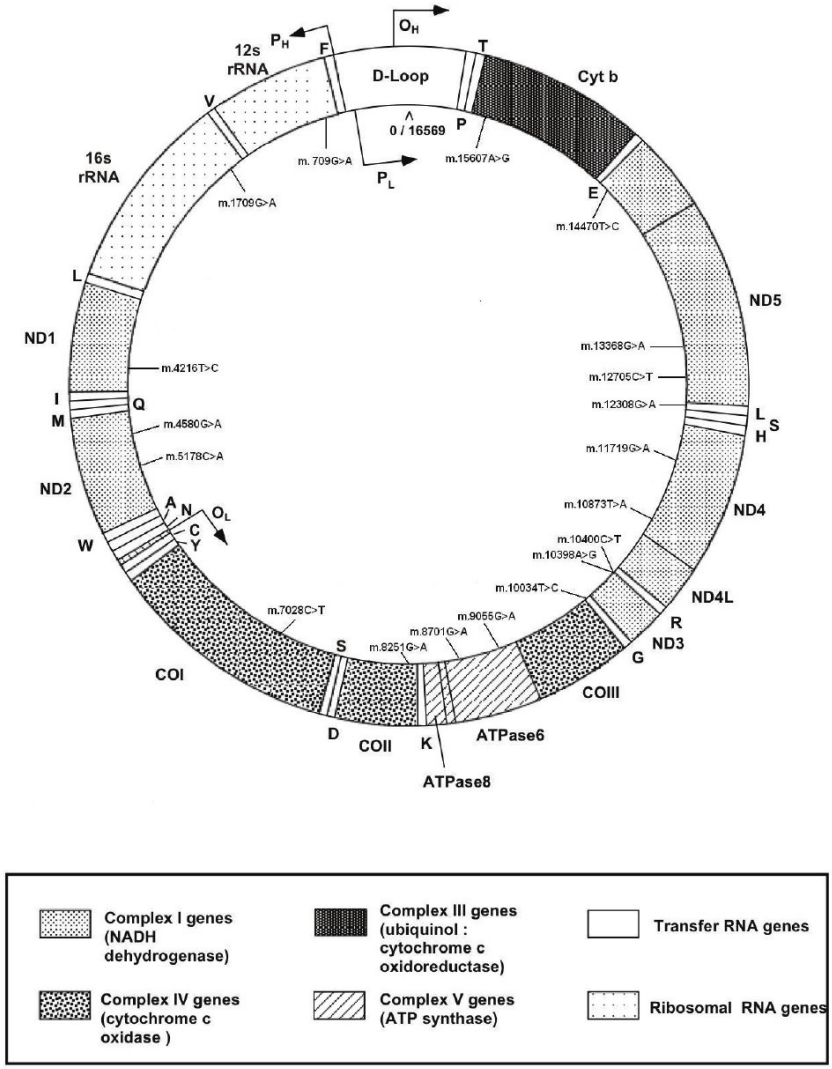
3.2. SUPPLEMENTARY MATERIAL

	SNP ID	m.709G>A	m.1719G>A	m.4216T>C	m.4580G>A	m.5178C>A	m.7028T>C	m.8251G>A	m.8701A>G	m.9055G>A	m.10034T>C	m.10398A>G	m.10400C>T	m.10873T>C	m.11719G>A	m.12308A>G	m.12705C>T	m.13368G>A	m.14470T>C	m.15607A>G	Haplogroup freq. (%)
	SNP type	ncod	ncod	p.Y304H	syn	p.I237M	syn	syn	p.T59A	p.A177T	ncod	p.T114A	syn	syn	syn	ncod	syn	syn	syn	syn	
Haplogroup	H						**C**								**G**		C				16.4
	R0				**G**		**T**								**G**		C				13.6
	V				**A**		T								**G**						0.0
	J	**G**	G	**C**								**G**			A		**C**	**G**			14.0
	T	**A**		**C**			T					**A**			A		C	**A**		**G**	9.1
	U						T		**A**	**G**				T		**G**	C			A	20.6
	K								A	**A**						**G**	C			A	6.5
	R*					C			**A**					T	A	A	**C**			**A**	2.1
	N*		G					**G**	**A**		T					A	**T**		**T**		1.2
	N1		**A**					**G**	**A**		**T**	**A**					T				0.9
	N1e'I		**A**						A		**T**						T				2.1
	I		**A**					**A**	A		**C**	G					T				3.0
	W	**A**	**G**					**A**	A			A				A	T				2.6
	X								A								T		**C**		1.6
	L								**G**		T	G	**C**	C			T				0.7
	M*					**C**			G		T		**T**	**C**			T				4.0
	D					**A**							**T**	**C**							1.6
SNP freq. (%)	Controls	14.5	7.0	23.4	0.0	1.6	83.2	7.5	6.3	7.0	3.0	29.0	5.6	6.3	70.1	27.1	17.5	9.3	1.9	9.3	
	ICBD cases	19.6	7.2	24.5	0.3	1.5	80.9	7.9	5.1	7.9	2.6	28.0	4.6	5.2	69.7	25.5	16.5	11.3	2.1	11.0	
	ISG cases	21.4	6.9	25.5	0.4	1.2	82.3	8.4	4.9	7.3	2.2	27.3	4.3	4.9	70.5	24.6	17.1	12.0	2.4	11.6	
ICBD cases vs. controls	$P_{unadjusted}$	**0.038**	0.980	0.714	0.641	0.963	0.299	0.914	0.474	0.696	0.832	0.782	0.548	0.550	0.952	0.631	0.778	0.365	0.945	0.458	
	OR[95% CI]	1.46[1.03-2.02]																			
	$P_{adjusted}$	0.053	0.997	0.665	0.972	0.859	0.296	0.667	0.436	0.635	0.843	0.836	0.514	0.517	0.959	0.562	0.862	0.394	0.856	0.503	
	OR[95% CI]	1.40[1.00-1.97]																			

ISG cases	$P_{unadjusted}$	**0.007**	0.924	0.459	0.537	0.807	0.690	0.702	0.442	0.905	0.595	0.683	0.439	0.446	0.784	0.449	0.972	0.213	0.704	0.292
vs. controls	OR[95% CI]	1.61[1.14-2.27]																		
	$P_{adjusted}$	**0.013**	0.890	0.505	0.971	0.759	0.585	0.536	0.395	0.839	0.579	0.671	0.413	0.437	0.867	0.397	0.981	0.238	0.585	0.330
	OR[95% CI]	1.56[1.10-2.21]																		

R*: all lineages inside haplogroup R except for haplogroups JT and UK; N*: all lineages inside haplogroup N except for haplogroups N1, I, W and X; M*: macro-haplogroup M except for haplogroup D.

Figure 1. Characterization and association of the investigated mitochondrial markers and haplogroups. Each haplogroup was determined by the combination of bolded alleles, and the alleles not bolded aided in the phylogenetic assignment. The polymorphisms are named after their base pair position and alleles. The type of the variant is indicated as ncod for non-coding SNPs, syn for synonymous SNPs, and the amino acid substitution is shown for non-synonymous SNPs. The haplogroup frequencies in controls and the SNP frequencies of the derived allele (second allele in the SNP name) are indicated. The results of mitochondrial SNP association testing with BD risk using the ICBD cases or the cases fulfilling ISG criteria are shown. Both unadjusted and adjusted (for gender) *P*-values are presented and significant *P*-values (<0.05) are highlighted in bold. Odds ratios (OR) and 95% confidence intervals (CI) are shown only for allele A of m.709G>A. This figure has been shown previously in the published paper (chapter 3.1, page 68), but given its poor printing quality it is shown again in this section with better quality.



Supplementary Figure 1. Genomic localization of the investigated markers within the human mitochondrial DNA molecule. The genetic location of mtDNA markers genotyped in the present study is indicated in the inner circle. H and L stand for heavy and light strands, respectively, given the asymmetric distribution of G and C nucleotides, with H being the G-rich strand. The seven complex I subunits (ND1, 2, 3, 4L, 4, 5 and 6), one complex III subunit (Cyt b), three complex IV subunits (COI, COII, and COIII), two complex V subunits (ATPases 6 and 8), two ribosomal RNAs (12S and 16S rRNAs), 22 tRNAs and D-loop regions are shown. Gene products encoded by the L-strand are shown in the inner circle (one letter code) while the products of the H-strand are shown in the outer circle. Arrows indicate the locations of promoters PL and PH for the transcription and replication origin O. Adapted from MITOMAP [http://www.mitomap.org]. This figure was not included in the published paper but it is deemed useful to ease the understanding of the work presented here.

CHAPTER 4

Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients

(Xavier *et al.,* Arthritis Rheum. 2012)

4.1 MAIN MANUSCRIPT

ARTHRITIS & RHEUMATISM
Vol. 64, No. 8, August 2012, pp 2761–2772
DOI 10.1002/art.34437

Association Study of *IL10* and *IL23R–IL12RB2* in Iranian Patients With Behçet's Disease

Joana M. Xavier,[1] Farhad Shahram,[2] Fereydoun Davatchi,[2] Alexandra Rosa,[3] Jorge Crespo,[4] Bahar Sadeghi Abdollahi,[2] Abdolhadi Nadji,[2] Gorete Jesus,[5] Filipe Barcelos,[6] José Vaz Patto,[6] Niloofar Mojarad Shafiee,[2] Fahmida Ghaderibarim,[2] and Sofia A. Oliveira[1]

Objective. **Independent replication of the findings from genome-wide association studies (GWAS) remains the gold standard for results validation. Our aim was to test the association of Behçet's disease (BD) with the interleukin-10 gene (*IL10*) and the IL-23 receptor–IL-12 receptor β2 (*IL23R–IL12RB2*) locus, each of which has been previously identified as a risk factor for BD in 2 different GWAS.**

Methods. **Six haplotype-tagging single-nucleotide polymorphisms (SNPs) in *IL10* and 42 in *IL23R–IL12RB2* were genotyped in 973 Iranian patients with BD and 637 non-BD controls. Population stratification was assessed using a panel of 86 ancestry-informative markers.**

Results. **Subtle evidence of population stratification was found in our data set. In *IL10*, rs1518111 was nominally associated with BD before and after adjustment for population stratification (odds ratio [OR] for T allele 1.20, 95% confidence interval [95% CI] 1.02–1.40, unadjusted P [P_{unadj}] = 2.53×10^{-2}; adjusted P [P_{adj}] = 1.43×10^{-2}), and rs1554286 demonstrated a trend toward association ($P_{unadj} = 6.14 \times 10^{-2}$; $P_{adj} = 3.21 \times 10^{-2}$). Six SNPs in *IL23R–IL12RB2* were found to be associated with BD after Bonferroni correction for multiple testing, the most significant of which were rs17375018 (OR for G allele 1.51, 95% CI 1.27–1.78, $P_{unadj} = 1.93 \times 10^{-6}$), rs7517847 (OR for T allele 1.48, 95% CI 1.26–1.74, $P_{unadj} = 1.23 \times 10^{-6}$), and rs924080 (OR for T allele 1.29, 95% CI 1.20–1.39, $P = 1.78 \times 10^{-5}$). SNPs rs10489629, rs1343151, and rs1495965 were also significantly associated with BD in all tests performed. Results of meta-analyses of our data combined with data from other populations further confirmed the role of rs1518111, rs17375018, rs7517847, and rs924080 in the risk of BD, but no epistatic interactions between *IL10* and *IL23R–IL12RB2* were detected. Results of imputation analysis highlighted the importance of *IL23R* regulatory regions in the susceptibility to BD.**

Conclusion. **These findings independently confirm, extend, and refine the association of BD with *IL10* and *IL23R–IL12RB2*. These associations warrant further validation and investigation in patients with BD, as they may have implications for the development of novel therapies (e.g., immunosuppressive therapy targeted at IL-23p19).**

Behçet's disease (BD) is a chronic inflammatory disorder classified as a vasculitis, and may involve several organs, such as the skin, mucocutaneous membranes (oral and genital aphthae), eyes, joints, lungs, and the gastrointestinal and central nervous systems. Although the etiology of BD remains unclear, an excessive immune or inflammatory response triggered by the exposure to an unknown infectious or environmental agent in genetically predisposed subjects is a likely pathogenic mechanism. Genome-wide association stud-

Supported by the Research Deputy of Tehran University of Medical Sciences (grant 132/714), the Portuguese Fundação para a Ciência e a Tecnologia (grant PTDC/SAU-GMG/098937/2008 and doctoral fellowship grant SFRH/BD/43895/2008 to Ms Xavier), and the Portuguese Instituto do Emprego e Formação Profissional (fellowship to Ms Xavier).

[1]Joana M. Xavier, MSc, Sofia A. Oliveira, PhD: Instituto de Medicina Molecular, Universidade de Lisboa, Lisbon, Portugal, and Instituto Gulbenkian de Ciência, Oeiras, Portugal; [2]Farhad Shahram, MD, Fereydoun Davatchi, MD, Bahar Sadeghi Abdollahi, MD, MPH, Abdolhadi Nadji, MD, Niloofar Mojarad Shafiee, MSc, Fahmida Ghaderibarim, MD: Tehran University of Medical Sciences, Tehran, Iran; [3]Alexandra Rosa, PhD: Universidade da Madeira, Funchal, Portugal; [4]Jorge Crespo, MD: Hospitais da Universidade de Coimbra, Coimbra, Portugal; [5]Gorete Jesus, MD: Hospital Infante D. Pedro, Aveiro, Portugal; [6]Filipe Barcelos, MD, José Vaz Patto, MD: Instituto Português de Reumatologia, Lisbon, Portugal.

Address correspondence to Sofia A. Oliveira, PhD, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Avenida Professor Egas Moniz, Edifício Egas Moniz, 1649-028 Lisbon, Portugal. E-mail: aaoliveira@fm.ul.pt.

Submitted for publication March 23, 2011; accepted in revised form February 9, 2012.

ies (GWAS) are currently the preferred method for uncovering novel genetic risk variants, and their results have brought novel insights into the biologic and genetic underpinnings of many complex disorders.

Two recent GWAS have confirmed the well-established *HLA–B51* association with the risk of BD and identified independent associations with the *HLA–A* region, the interleukin-10 gene (*IL10*), and the IL-23 receptor–IL-12 receptor $\beta 2$ (*IL23R–IL12RB2*) locus (1,2). Replication of the association findings in independent samples remains the sine qua non for validating susceptibility genes. In this study, we investigated the non–major histocompatibility complex (MHC) genetic component of BD by testing the association of 48 single-nucleotide polymorphisms (SNPs) in *IL10* and *IL23R–IL12RB2* with BD in a large data set of Iranian cases and controls, controlling for population stratification and testing for gene–gene interactions.

PATIENTS AND METHODS

Study subjects. The study data set comprised 973 patients with BD and 637 control subjects who were enrolled in the Behçet's Disease Clinic at the Rheumatology Research Center of Shariati Hospital (Tehran University of Medical Sciences, Tehran, Iran). The clinical and demographic features of the study participants were obtained by medical interview at the time of blood sampling and inspection of medical records. BD cases were selected as consecutive patients, and the diagnosis of BD was made according to the revised International Criteria for Behçet's Disease (3). Patients who were ≥60 years of age at the onset of BD were excluded.

Controls were evaluated using the same evaluation procedures as used for the cases, and were selected as a control if they did not have a diagnosis of BD, any other rheumatologic or autoimmune disorder, or oral aphthosis. The participants self-reported their ancestry as West Eurasian Caucasoid or East/Central Asian (Turk).

This study received ethics approval from the ethics committee at the Tehran University of Medical Sciences in Iran. All participants were informed of the study and provided their written informed consent to participate. The study was conducted in accordance with the Declaration of Helsinki.

SNP selection. A total of 50 haplotype-tagging SNPs (tag SNPs) in *IL10* and *IL23R–IL12RB2,* including the SNPs showing the most significant association with BD in the 2 previously published GWAS (rs1518111 and rs1554286 in *IL10*, and rs924080 and rs1495965 in *IL23R–IL12RB2*) (1,2), were selected for genotyping. In *IL10*, rs1554286 was selected for study because the other SNPs that showed stronger associations in the Japanese GWAS (rs1800871, rs1800872, and rs1518111) were in very strong linkage disequilibrium (LD) ($0.933 \leq r^2 \leq 1$) in individuals from the CEPH HapMap population (Utah residents with ancestry from northern and western Europe; CEU).

Tag SNPs in *IL10* (chromosome 1; 205006500–205017000) and in the *IL23R–IL12RB2* locus (chromosome 1; 67400757–67641236) were identified in Haploview, version 4.1 (4), using genotypes of 30 European (CEU) family trios downloaded from the HapMap database (release 24/phase II, November 2008, NCBI B36 assembly, dbSNP b126). In this analysis, the following options were utilized: pairwise mode of analysis, LD values of $r^2 \geq 0.90$ in *IL10* and $r^2 \geq 0.80$ in *IL23R–IL12RB2*, and minor allele frequency (MAF) values of ≥0.10. To control for population stratification, we selected the 100 most-informative biallelic ancestry-informative markers (AIMs) (for a full list, see Supplementary Table 1, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1529-0131), i.e., those markers that could distinguish the northwestern from southeastern European ancestries of European Americans (5). These AIMs were used to correct for population substructure in our data set (for a detailed description of this approach, see http://genepath.med.harvard.edu/~reich/EUROSNP.htm).

Genotyping. Genomic DNA was extracted from samples of whole blood using a salting-out procedure. SNPs were genotyped using a Sequenom iPlex assay (primer extension of multiplex products with detection by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) following the manufacturer's protocol, with detection in a Sequenom MassArray K2 platform. The primer sequences (available from the corresponding author upon request) were designed using Sequenom MassARRAY Assay Design software (version 3.0). All of the genotyping was performed in the Genomics Unit of the Instituto Gulbenkian de Ciência in Portugal. Six AIMs and 2 SNPs in *IL23R–IL12RB2* failed in the genotyping assay design phase, and were therefore excluded from further analyses.

Extensive quality control was performed with the use of 8 HapMap controls of diverse ethnic affiliation (http://hapmap.ncbi.nlm.nih.gov/), along with sample duplication within and across plates, a Mendelian inheritance check in 3 large pedigrees, calculation of Hardy-Weinberg equilibrium (HWE) ($P > 0.01$) in the control group, and a minimum call rate of 95% for each SNP. Genotype determinations were performed with the investigators blinded to the affection status (i.e., case or control). Eight AIMs and 2 SNPs in *IL23R–IL12RB2* (rs10889668 and rs4655536) were excluded on the basis of quality control issues (e.g., genotyping errors, out of HWE in controls).

Population stratification assessment. Principal components analysis (PCA) and F_{ST} statistics were computed for the AIMs data, carried out using the EigenSoft program, version 3.0 (6), with default parameters. F_{ST} values indicate which proportion of the total genetic variance is attributable to the differences between populations and to the differences between individuals of the same population. This measure of genetic structure varies between 0 (no genetic differentiation of population units) and 1 (complete differentiation of population units). We ran PCA on 2 data sets: 1) the complete Iranian data set (973 patients and 637 controls), and 2) the Iranian samples plus 525 HapMap individuals from 4 populations (CEU n = 174, the Yoruba people of Ibadan, Nigeria [YRI] n = 176, individuals from Beijing, China [CHB] n = 86, and individuals from Tokyo, Japan [JPT] n = 89). The genotypes for these 525 HapMap individuals at the studied AIMs were downloaded from the International HapMap Proj-

ect web site (release 27; http://hapmap.ncbi.nlm.nih.gov/biomart/martview/). Genotypes were unavailable for AIMs 34 and 83 in all HapMap samples, and for AIMs 1 and 24 in the CHB and JPT populations.

Association testing. Student's unpaired *t*-tests and chi-square tests were used to compare quantitative clinical data and qualitative demographic data between BD cases and controls. Chi-square tests for HWE in controls, crude allelic association tests (results expressed as unadjusted P [P_{unadj}] values), and LD plot analyses were performed using Haploview, version 4.1 (4). Tests for allelic association adjusted for population stratification (results expressed as adjusted P [P_{adj}] values) were performed with EigenStrat (7). Conditional analyses based on the most significantly associated SNP (rs7517847) (results expressed as conditional P [P_{cond}] values) were performed with logistic regression using the SNPassoc version 1.4-9 package (8), implemented in R freeware (http://cran.r-project.org/). Odds ratios (ORs) and their associated 95% confidence intervals (95% CIs) were calculated for SNPs showing significant associations with BD. Results were considered significant at levels below the Bonferroni correction level ($P < 1.04 \times 10^{-3}$ for the 48 genotyped SNPs and $P < 4.76 \times 10^{-4}$ for the 105 imputed SNPs). Power analyses were performed using CaTS (9) with a multiplicative genetic model.

Meta-analyses. Fixed-effects (Mantel-Haenszel) meta-analyses were performed using the rmeta package in R.2.7.2, and Plink version 1.04 (10). Two methods were used to estimate between-study heterogeneity: the chi-square–based Cochran's Q statistic (with heterogeneity considered significant at P values [P_{het}] less than 0.10), and the I^2 metric ($I^2 <$ 25 = no heterogeneity; $25 \leq I^2 < 50$ = moderate heterogeneity; $50 \leq I^2 < 75$ = large heterogeneity; $75 \leq I^2 < 100$ = extreme heterogeneity). Forest plots were drawn to visualize the overall effect.

Gene–gene interactions. Epistasis between *IL10* and *IL23R–IL12RB2* was investigated using the multifactor-dimensionality reduction (MDR) method (http://www.epistasis.org) (11). Missing genotypes were first imputed for each SNP from the Plink database (10), using genotypes from individuals in the CEU HapMap population as a reference, and MDR was run using several seeds.

Genotype imputation. Imputation of genotypes is a statistical tool that uses the correlation between markers present in a reference data set (such as HapMap) to predict the unobserved genotypes present in an experimental sample. An imputation algorithm uses both the dense information from the reference sample and the less-dense genotype information from the experimental sample to infer genotypes at SNPs not directly genotyped in the study.

Ungenotyped SNPs in chromosome 1 were imputed with the Plink program, version 1.04 (10), using HapMap data (release 22; 173,899 SNPs in chromosome 1, 60 CEU founders). For every imputed SNP, Plink provides an information content metric, referred to as SNP INFO, with values ranging from 0 to 1 (although it can be greater than 1 occasionally). A higher INFO value generally means a better SNP imputation. All imputed SNPs with MAF values less than 0.1 in controls and with INFO values less than 0.800 were excluded. For each SNP that has been genotyped, Plink calculates the concordance rate, which is determined by comparing the experimentally determined genotypes with those imputed from surrounding genotyped markers.

RESULTS

Characterization of the data set. The general characteristics of the subjects in our Iranian data set are shown in Table 1. The distribution of clinical symptoms in the patients with BD (e.g., 98.7% with oral aphthosis, 62.3% with genital aphthosis, 55.1% with skin lesions, and 59.8% with ocular lesions) is in line with what has been observed in larger data sets (12), suggesting that this is a representative patient group. Since the male-to-female ratio was not significantly different between the control and patient groups (54.6% male versus 52.5% male; $P = 0.406$) and there is no evidence that genetic susceptibility to BD varies depending on sex, the effect of this variable did not need to be taken into account in association tests. To minimize misclassification biases, we chose control subjects who were slightly older than the cases (mean ± SD age at examination 42.0 ± 11.6 years in controls and 39.1 ± 11.0 years in cases). Again, as with sex, age is not a risk factor for BD, and therefore this variable did not need to be taken into account in association analyses.

Population stratification. Population stratification (differences in allele frequency between cases and controls that are attributable to ancestry differences or to ancestry-specific disease risks) has been a major confounding factor and a source of inconsistency and type I errors in association studies that do not generate dense genotype data. To assess whether a hidden population substructure was present in our Iranian sample, we genotyped 86 AIMs in our full data set (Supplementary Table 1, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1529-0131). These 86 AIMs are among the top 100 most-informative markers found to reliably discern the major axes of ancestry variation in European Americans (5). PCA is the most commonly used tool to ascertain populations based on genetic structure, regardless of the self-reported data. Figure 1A shows the top 2 principal components of genetic variation identified by PCA in the Iranian data set and in the HapMap reference populations (174 individuals in CEU, 86 in CHB, 89 in JPT, and 176 in YRI). Eigenvectors 1–3 were statistically significantly different between populations ($P < 0.05$), contributing 5.8%, 3.9%, and 2.3%, respectively, to the total variation in ancestry and representing a total of 11.9% of the cumulative variation.

Even though genotypes at the marker previously

Table 1. General characteristics of the study sample of Iranian patients with Behçet's disease (BD) and non-BD control subjects

Characteristic	Controls	BD cases
No. of subjects	637	973
Sex, no. (%) male	348 (54.6)	511 (52.5)
Age at examination, mean ± SD years	42.0 ± 11.6	39.1 ± 11.0
Age at diagnosis, mean ± SD years	–	32.2 ± 9.2
Self-reported ancestry, no./total assessed (%)		
West Eurasian Caucasoid	475/634 (74.9)	564/969 (58.2)
East/Central Asian (Turk)	159/634 (25.1)	405/969 (41.8)
Clinical symptom, no./total assessed (%)		
Oral aphthosis		960/973 (98.7)
Genital aphthosis		606/973 (62.3)
Skin lesions		536/973 (55.1)
Pseudofolliculitis		408/536 (76.1)
Erythema nodosum		210/536 (39.2)
Skin aphthosis		29/973 (3.04)
Ophthalmologic manifestations		582/973 (59.8)
Anterior uveitis		419/582 (72.0)
Posterior uveitis		494/582 (84.9)
Retinal vasculitis		342/582 (58.8)
Joint manifestations		300/973 (30.8)
Arthralgia		138/300 (46.0)
Arthritis		188/300 (62.7)
Ankylosing spondylitis		17/300 (5.7)
Neurologic manifestations		62/973 (6.4)
Vascular involvement		51/973 (5.2)
Gastrointestinal manifestations		39/973 (4.0)
Epididymitis		22/511 (4.3)
Cardiac involvement		6/973 (0.6)
Pleuropulmonary involvement		6/973 (0.6)
Pathergy phenomenon		439/957 (45.9)
Family history of BD, no./total assessed (%)		86/933 (9.2)

reported to be the most-informative AIM (AIM 83) (5) were not available for all HapMap samples, we observed that HapMap populations tended to cluster into distinct groups of genetic variation (Figure 1A), with the exception of the CHB and JPT samples, which showed overlap of genotypes. This suggests that these 86 AIMs have good power to discriminate samples in terms of discerning European, African, and Asian ancestries. Iranian subjects are in closer proximity to the CEU cluster of subjects than to any other HapMap group, and the distribution of genetic variation overlapped in BD cases and controls (Figures 1A and B), with no major outliers. These findings are in agreement with the F_{ST} pairwise statistics, which showed similar patterns of genetic variance between HapMap populations and between BD cases and controls (as shown in Supplementary Table 2, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1529-0131).

Population stratification was not readily observed in the PCA plot without the HapMap reference samples (Figure 1B), and an F_{ST} value of <0.0005 for the genetic variance between BD cases and controls suggests that there was genetic homogeneity between the patients with BD and the control subjects in our data set. Analysis of variance for population differences along the eigenvectors demonstrated that BD cases and controls did not differ with regard to the first 2 principal

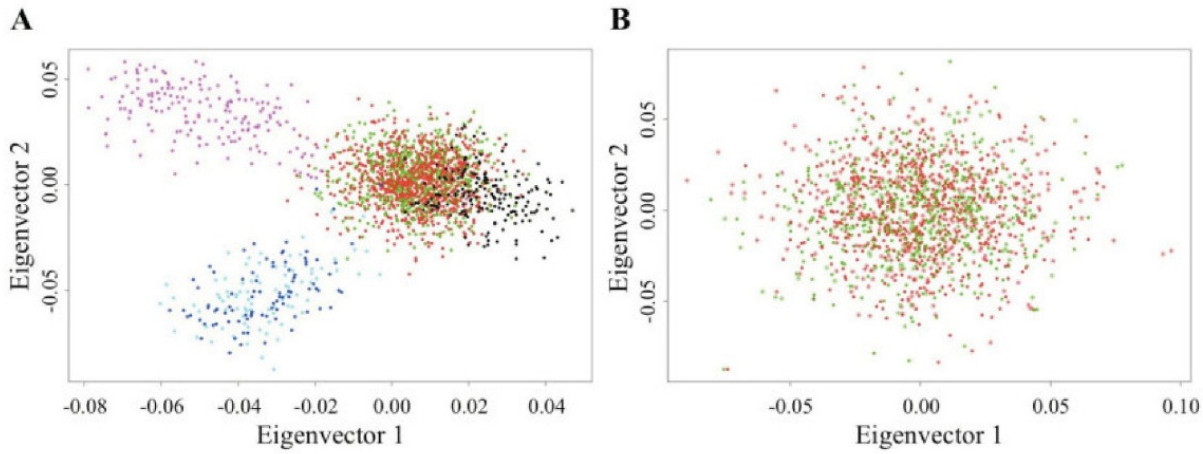


Figure 1. Top 2 axes of genetic variation, determined as eigenvectors in principal components analysis of genotypes at 86 ancestry-informative markers, in the 973 Iranian patients with Behçet's disease (BD) and 637 controls. In **A,** individual BD case and control samples are shown along with HapMap reference samples (CEPH population of Utah residents with ancestry from northern and western Europe [CEU] n = 174, Yoruba people of Ibadan, Nigeria [YRI] n = 176, individuals from Beijing, China [CHB] n = 86, and individuals from Tokyo, Japan [JPT] n = 89). In **B,** only BD case and control samples are shown, without HapMap reference samples. BD cases are shown in red, controls in green, CEU in black, YRI in pink, CHB in light blue, and JPT in dark blue. In **B,** dots and asterisks indicate individuals of self-reported West Eurasian Caucasoid and East/Central Asian ancestries, respectively.

components nor did they differ in the overall PCA test findings ($P = 0.242$). However, there was a significant effect of affection status (case versus control) on eigenvectors 3, 6, 13, 19, 44, and 74 (P values equal to 0.021, 0.020, 0.004, 0.037, 0.033, and 0.036, respectively). Given that no significant difference between cases and controls was observed for any of the remaining principal components and that eigenvector 13 showed the most significant difference, the first 13 principal components were used as covariates for ancestry adjustment in association tests.

In summary, we have further validated the capacity of the selected AIMs to resolve differences in intercontinental ancestry, and our results suggest that subtle ancestry differences can be detected in our data set. Thus, adjustment for the top 13 eigenvectors is an important requisite step in association tests.

Association of SNPs with BD. Six tag SNPs in *IL10* and 40 tag SNPs in *IL23R–IL12RB2* were successfully genotyped in our data set (SNPs 1–6 and 9–48; listed in Supplementary Table 3, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1529-0131). The plots of pairwise LD of the genotyped SNPs in the Iranian samples (Figure 2A) suggested that these markers were generally not correlated and, therefore, captured existing genetic diversity in these loci. Furthermore, the LD patterns were very similar in the Iranian samples and the CEU samples (Figures 2A and B), which strongly supports the use of the HapMap CEU data set for selecting tag SNPs and as a reference to perform imputation in the Iranian population.

In *IL10*, a trend for association with BD was observed for the SNPs rs1518111 (OR for T allele 1.20, 95% CI 1.02–1.40, $P_{unadj} = 2.53 \times 10^{-2}$ and $P_{adj} = 1.43 \times 10^{-2}$) and rs1554286 ($P_{unadj} = 6.14 \times 10^{-2}$ and $P_{adj} = 3.21 \times 10^{-2}$) (Table 2). Our data set had an 89–100% power to detect an association for these 2 polymorphisms (disease allele frequency 0.220–0.332) with a magnitude of association (OR 1.45–1.61) similar

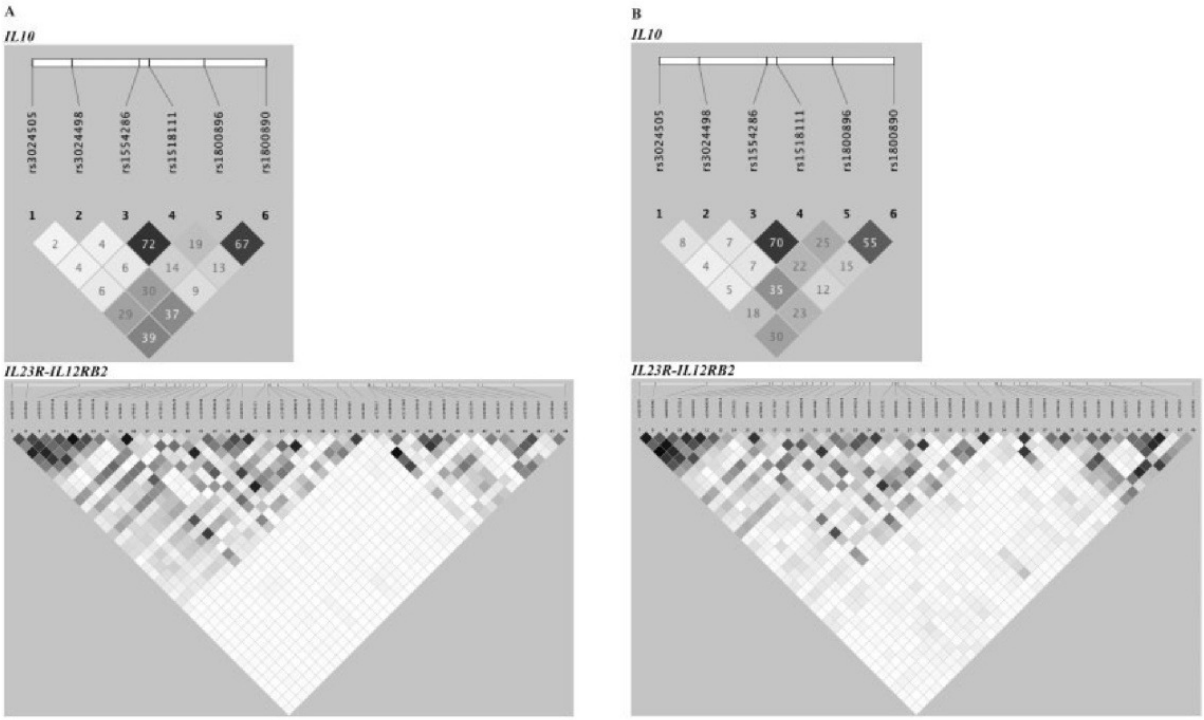


Figure 2. Linkage disequilibrium plots in the Iranian data set **(A)** and in 30 CEU (CEPH population of Utah residents with ancestry from northern and western Europe) HapMap trios **(B)** for the genotyped single-nucleotide polymorphisms (SNPs) in *IL10* and *IL23R–IL12RB2*. These plots were obtained in Haploview using the r^2 color scheme (white indicating $r^2 = 0$ and black indicating $r^2 = 1$, with different shades of grey indicating $0 < r^2 < 1$). In addition, values in the *IL10* plots indicate r^2 values for pairwise comparisons between the SNPs.

Table 2. Single-nucleotide polymorphisms (SNPs) associated with the risk of Behçet's disease

Gene or locus, SNP identifier	SNP	OR (95% CI)*	*P* value†		
			P_{unadj}	P_{adj}	P_{cond}
IL10					
SNP 1	rs3024505		4.83×10^{-1}	4.99×10^{-1}	
SNP 2	rs3024498		8.84×10^{-1}	9.62×10^{-1}	
SNP 3	rs1554286		6.14×10^{-2}	3.21×10^{-2}	
SNP 4	rs1518111		2.53×10^{-2}	1.43×10^{-2}	
SNP 5	rs1800896		8.24×10^{-1}	9.25×10^{-1}	
SNP 6	rs1800890		6.72×10^{-1}	7.99×10^{-1}	
IL23R–IL12RB2					
SNP 7	rs4655679	0.69 (0.58–0.81)	1.54×10^{-5}‡	2.06×10^{-5}‡	4.25×10^{-2}
SNP 8	rs9729046		2.90×10^{-3}	6.96×10^{-3}	
SNP 9	rs6683039		6.04×10^{-1}	8.25×10^{-1}	
SNP 10	rs17375018	1.51 (1.27–1.78)	1.93×10^{-6}‡	2.31×10^{-6}‡	1.73×10^{-2}
SNP 11	rs6656929		1.70×10^{-3}	2.84×10^{-3}	
SNP 12	rs10489630		2.00×10^{-3}	2.58×10^{-3}	
SNP 13	rs11209018		1.76×10^{-1}	2.25×10^{-1}	
SNP 14	rs7539625		3.18×10^{-2}	2.90×10^{-1}	
SNP 15	rs790631		1.10×10^{-2}	1.29×10^{-2}	
SNP 16	rs790633		1.64×10^{-2}	2.01×10^{-2}	
SNP 17	rs7517847	1.48 (1.26–1.74)	1.23×10^{-6}‡	1.96×10^{-6}‡	Referent
SNP 18	rs7530511		1.20×10^{-3}	1.23×10^{-3}	
SNP 19	rs10489629	0.77 (0.66–0.9)	7.00×10^{-4}‡	1.07×10^{-3}‡	9.48×10^{-1}
SNP 20	rs4655692		1.70×10^{-2}	1.43×10^{-2}	
SNP 21	rs12030948		9.52×10^{-1}	9.16×10^{-1}	
SNP 22	rs10489628		3.23×10^{-2}	3.70×10^{-2}	
SNP 23	rs10789229		7.81×10^{-1}	8.78×10^{-1}	
SNP 24	rs6682033		6.90×10^{-3}	7.99×10^{-3}	
SNP 25	rs1343151	0.72 (0.61–0.86)	2.00×10^{-4}‡	1.87×10^{-4}‡	2.26×10^{-1}
SNP 26	rs6693831		9.90×10^{-3}	8.06×10^{-3}	
SNP 27	rs11465817		2.59×10^{-1}	3.05×10^{-1}	
SNP 28	rs10889675		6.64×10^{-1}	6.43×10^{-1}	
SNP 29	rs10889677		2.11×10^{-1}	2.68×10^{-1}	
SNP 30	rs11209030		4.40×10^{-3}	3.41×10^{-3}	
SNP 31	rs17303361		2.38×10^{-1}	2.37×10^{-1}	
SNP 32	rs1495965	0.75 (0.65–0.87)	1.00×10^{-4}‡	1.26×10^{-4}‡	2.65×10^{-2}
SNP 33	rs924080	0.72 (0.62–0.84)	1.78×10^{-5}‡	2.05×10^{-5}‡	1.56×10^{-3}
SNP 34	rs7539817		9.44×10^{-1}	9.74×10^{-1}	
SNP 35	rs10889680		5.24×10^{-1}	5.43×10^{-1}	
SNP 36	rs12131065		9.13×10^{-1}	9.43×10^{-1}	
SNP 37	rs11209045		5.79×10^{-1}	6.11×10^{-1}	
SNP 38	rs3790558		9.43×10^{-1}	9.75×10^{-1}	
SNP 39	rs10489627		9.26×10^{-1}	8.17×10^{-1}	
SNP 40	rs1890741		9.17×10^{-1}	9.08×10^{-1}	
SNP 41	rs2201584		8.50×10^{-1}	8.21×10^{-1}	
SNP 42	rs6693065		7.31×10^{-1}	7.85×10^{-1}	
SNP 43	rs1995147		9.10×10^{-1}	8.28×10^{-1}	
SNP 44	rs3790565		9.94×10^{-1}	9.09×10^{-1}	
SNP 45	rs6679356		4.62×10^{-1}	4.18×10^{-1}	
SNP 46	rs3790567		4.37×10^{-1}	3.94×10^{-1}	
SNP 47	rs3790569		8.86×10^{-1}	8.79×10^{-1}	
SNP 48	rs2229546		8.76×10^{-1}	9.27×10^{-1}	

* The respective odds ratios (ORs) and 95% confidence intervals (95% CIs) were determined in the presence of the following specific SNP alleles: T allele of rs4655679, rs7517847, and rs1343151, A allele of rs1495965, G allele of rs17375018 and rs10489629, and C allele of rs924080.

† For each SNP tested, the crude allelic *P* value (unadjusted *P* [P_{unadj}]) and the *P* value adjusted for the top 13 eigenvectors of population genetic variation (P_{adj}), using EigenStrat, are indicated. For the SNPs showing a significant association with Behçet's disease in these tests, the *P* value from the association analysis conditioned on the most significantly associated SNP, rs7517847 (P_{cond}), is also indicated.

‡ *P* value shows a significant association with Behçet's disease at $P < 1.04 \times 10^{-3}$.

to that described previously (1,2). These associations in *IL10* did not withstand the conservative Bonferroni correction for multiple testing.

In *IL23R–IL12RB2*, 6 markers (SNPs 10, 17, 19, 25, 32, and 33 in Table 2) demonstrated a significant association with the risk of BD ($P < 1.04 \times 10^{-3}$) both before and after adjustment for population stratification (Table 2). Except for SNPs 32 and 33, which had an

r^2 value of 0.66 for LD in our data set, these 6 SNPs were not in strong LD ($r^2 < 0.40$) (Figure 2A). SNP 10 (rs17375018) and linked SNPs 32 (rs1495965) and 33 (rs924080) remained only nominally associated with BD after conditional logistic regression analysis based on the most significant polymorphism, rs7517847 (Table 2). Haplotype associations did not provide additional insights over single-marker associations (results not shown).

Meta-analysis and epistasis. A combined analysis of the genetic markers showing the most significant association with BD in our Iranian data set and the data from other populations in previous reports (1,2,13) revealed an increased risk of BD associated with the T allele of rs1518111 in *IL10* (OR for T allele 1.39, 95% CI 1.29–1.50, $P = 3.54 \times 10^{-23}$, $P_{het} = 0.138$, $I^2 = 36.5$), and an increased risk of BD associated with the G allele in rs17375018 (OR for G allele 1.44, 95% CI 1.30–1.60, $P = 5.86 \times 10^{-12}$, $P_{het} = 0.460$, $I^2 < 0.05$), the T allele in rs7517847 (OR for T allele 1.36, 95% CI 1.23–1.50, $P = 1.88 \times 10^{-9}$, $P_{het} = 0.285$, $I^2 = 20.4$), and the T allele in rs924080 (OR for T allele 1.29, 95% CI 1.20–1.39, $P = 9.77 \times 10^{-13}$, $P_{het} = 0.472$, $I^2 < 0.05$) in *IL23R–IL12RB2* (forest plots are presented in Figure 3). The consistency of the association of these SNPs in *IL10* and *IL23R–IL12RB2*, as revealed by this meta-analysis, further supports the role of these SNPs in the etiology of BD.

We then explored the occurrence of nonadditive gene–gene interactions between *IL10* and *IL23R–IL12RB2* in the risk of BD, using the MDR method, which is a nonparametric and genetic-model–free approach for identifying high-order gene–gene interactions in the absence of detectable independent main effects (11). We tested all 2- to 4-locus interaction models, but none of the best models included polymorphisms in both the *IL10* and *IL23R–IL12RB2* loci, and therefore we did not detect any evidence for epistasis among these genes.

Results of imputation analysis. Even though we investigated 40 SNPs in *IL23R–IL12RB2* (SNPs 9–48),

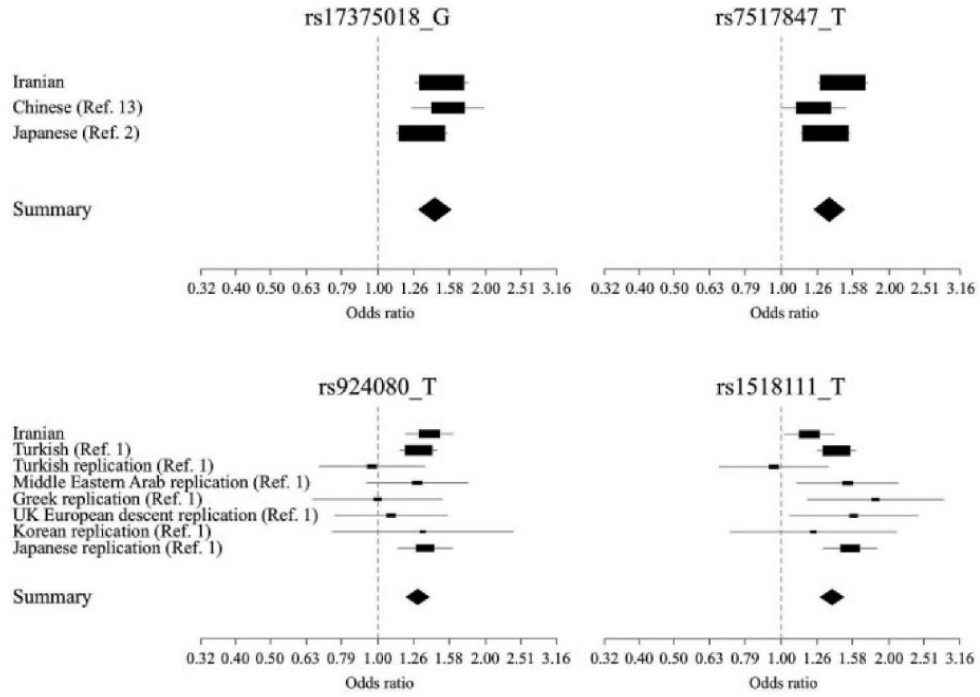


Figure 3. Forest plots for 4 markers, in *IL10* (rs1518111) and in *IL23R–IL12RB2* (rs17375018, rs7517847, and rs924080) significantly associated with Behçet's disease in the Iranian population compared with other populations. At each marker, the results for each population and for the overall meta-analysis are relative to the allele on the forward strand of the human genome reference sequence. Results are shown as the odds ratio, which is represented by a rectangle (size is proportional to the respective amount of data), and the 95% confidence interval, which is represented by bars. The pooled point estimate for the meta-analysis is represented by a diamond. The vertical broken line shows the no-effect point (odds ratio 1.00).

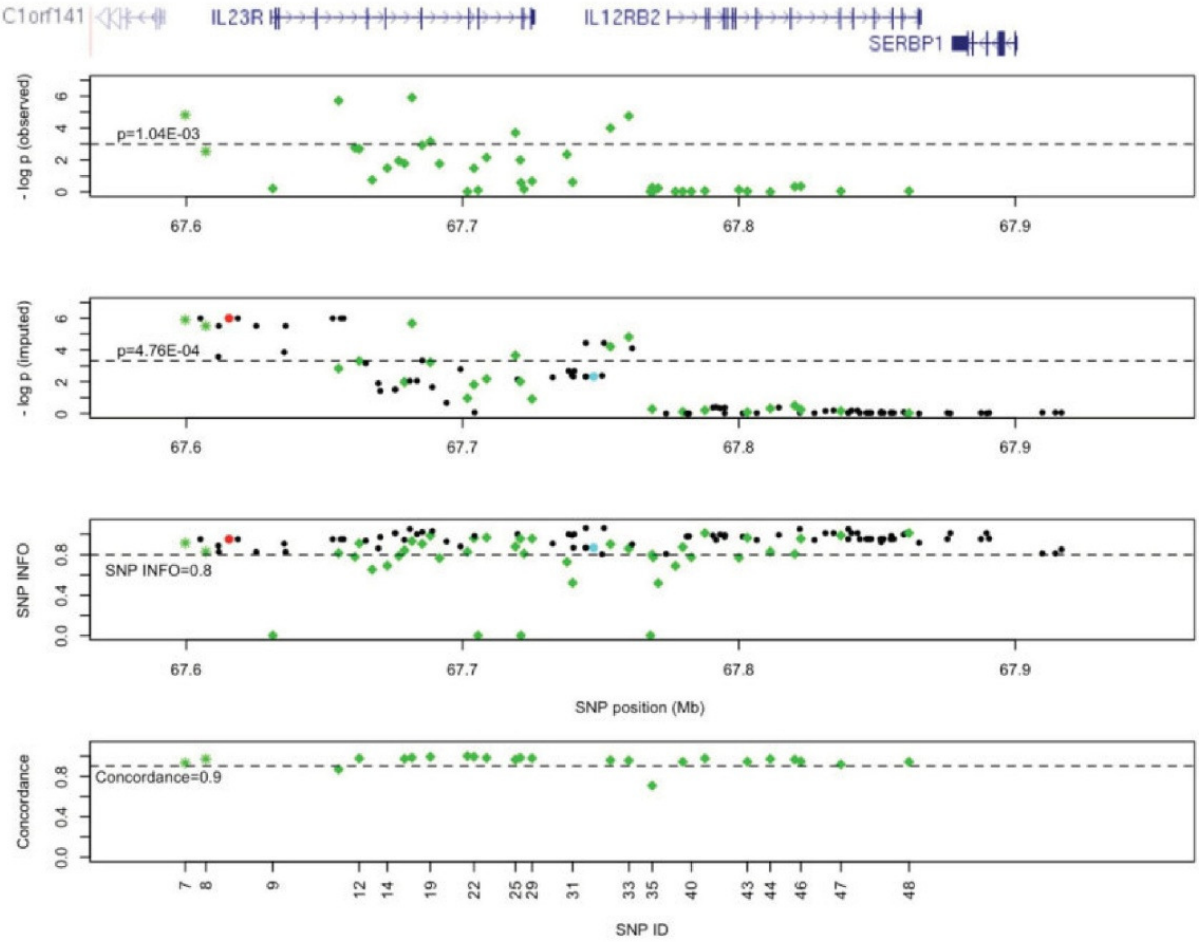


Figure 4. Results of association tests assessing the association of Behçet's disease with observed and imputed single-nucleotide polymorphisms (SNPs) in the *IL23R–IL12RB2* genomic region. The hatched bars above the plots represent the introns and exons of the indicated genes, relative to the genotyped and imputed polymorphisms. The top plot shows the negative logarithm of the *P* value (unadjusted allelic association test) for the 42 observed polymorphisms genotyped before (SNPs 9–48) and after (SNPs 7–8, represented by green asterisks) the imputation analyses. The second graph displays the negative logarithm of the *P* value (unadjusted allelic association test) for 105 SNPs in chromosome 1 imputed with high quality (SNPs with a minor allele frequency [MAF] ≥0.10 and SNP INFO ≥0.800), including SNPs that were genotyped. The third plot shows the SNP information content metric SNP INFO for each of the 105 SNPs. The bottom graph displays the rate of concordance of the observed and imputed genotypes (only for SNPs with imputed MAF ≥0.10 and SNP INFO ≥0.800). In all plots, the SNPs that had been initially genotyped (SNPs 9–48) are represented with green diamonds, and rs12119179 and rs10889657 are indicated by blue and red dots, respectively. Broken horizontal lines indicate the cutoff for the significance threshold. ID = identifier.

a large portion of its natural genetic variation was not assessed because it is a sizeable region with a high degree of genetic diversity. To evaluate the potential association of unobserved polymorphisms in this gene, we performed genotype imputation for SNPs in chromosome 1 using data from HapMap as well as the genotypes observed at the 40 fully genotyped markers. In the *IL23R–IL12RB2* genomic region, we obtained imputed genotypes meeting minimum quality standards (MAF values in controls ≥0.10 and SNP INFO values ≥0.800) for 105 SNPs, including 22 of the experimentally genotyped SNPs (the genotyped SNPs were imputed using the observed genotypes at the other SNPs, and a concordance rate of >85% between imputed and observed

genotypes was obtained for 21 of the 22 genotyped and imputed SNPs). Eighteen of the genotyped SNPs could not be accurately imputed, most likely because they are tag SNPs that are not captured by any combination of neighboring variants.

SNP rs12119179, which has been previously found to be associated with BD ($P = 2.70 \times 10^{-8}$) (2) but could not be genotyped in our data set, was not significantly associated with BD in our analyses using imputed genotypes ($P = 4.78 \times 10^{-3}$, SNP INFO = 0.868) (indicated by the light blue dot in Figure 4), when $P < 4.76 \times 10^{-4}$ was considered the significance threshold for association among the 105 imputed SNPs.

Furthermore, 21 imputed polymorphisms (including 4 of the experimentally genotyped SNPs) had an allelic association with the risk of BD at a significance threshold of $P < 4.76 \times 10^{-4}$ (Figure 4), the most significant of which was the rs10889657 SNP located 5′ of *IL23R* ($P = 1.02 \times 10^{-6}$, SNP INFO = 0.950) (indicated by the red dot in Figure 4).

Based on the observed pattern of LD between the markers (SNPs 9–48) in the HapMap CEU reference sample and the 17 significant imputed SNPs, we selected SNPs 7 and 8 for experimental validation of the imputation results. SNP rs10889657 was not selected since it is in very high LD with SNP 7 ($r^2 = 0.95$). We confirmed, by experimental genotyping, the association of SNPs 7 and 8 with BD ($P_{unadj} = 1.54 \times 10^{-5}$ and $P_{unadj} = 2.90 \times 10^{-3}$, respectively) (Table 2), and concordance rates between the imputed and observed genotypes were 0.932 and 0.971, respectively. The 21 significantly associated polymorphisms are located in *IL23R* and neighboring intergenic regions, but seem to exclude the downstream *IL12RB2* gene (Figure 4). The results of this imputation analysis therefore strongly support the idea that the observed associations in the *IL23R–IL12RB2* locus most likely originate from *IL23R* regulatory regions.

DISCUSSION

The findings from this case–control association study independently validated and extended the previous association of *IL10* and of the *IL23R–IL12RB2* locus with BD risk (1,2,13). Unlike what is typically observed in replication studies, our most significant findings are very consistent with those from previous reports (1,2,13), as reinforced by the meta-analyses. These genetic associations are also in line with our current understanding of the immunopathology of BD.

IL23R and *IL12RB2* encode for specific subunits of the IL-23 and IL-12 receptors, respectively, while the other subunit of these receptors is shared and encoded by *IL12RB1*. Moreover, IL-23 and IL-12 are related cytokines, since they share the p40 subunit (encoded by *IL12B*) but can be distinguished by their specific subunits (p35 for IL12, and p19, encoded by *IL23A*, for IL23) (14). Even though the IL-12 and IL-23 receptors are expressed in the same cell types (activated/memory T cells and natural killer cells) and the same JAK kinases and STAT proteins are activated by IL-12 and IL-23 (14), different biologic responses are expected since they occur under different stimuli, and there are differences in the DNA–STAT transcription factor complexes that are formed (e.g., the most prominent STAT induced by IL-12 is STAT-4, and that induced by IL-23 is STAT-3) (15).

IL-12 has a key role in promoting the differentiation of naive CD4+ T cells into mature Th1 effector cells expressing genes involved in host defense, whereas IL-23 is not required for the generation of Th17 cells, but is required for their maintenance (15). The hallmark cytokines produced by human Th17 cells are IL-17A and IL-17F, which are involved in the recruitment, activation, and migration of neutrophils, as well as IL-22. Th17 cells are potent inducers of tissue inflammation and appear to be involved in the clearance of pathogens not adequately handled by Th1 or Th2 cells alone, especially at epithelial/mucosal barriers. While our understanding of infectious and allergic diseases has greatly advanced with the discovery of Th1 and Th2 immune responses, more recently a growing list of autoimmune disorders has been associated with aberrant Th17 responses and the IL-23/IL-17 axis (16).

Early studies on the circulating CD4+ T cells and affected lesions of BD patients with active disease showed elevated levels of Th1 cytokines, such as interferon-γ (IFNγ) and IL-12, suggesting that a Th1-polarized immune response plays a major role in the BD pathogenic process (17–19). More recently, elevated levels of IL-23, IL-17, and IFNγ have been observed in BD patients with active uveitis (20), and the Th17:Th1 ratio was found to be significantly increased in patients with BD (21). Of note, different T cell subsets, namely Th17, can display plasticity (22), and in patients with BD, a minor proportion of T cells can simultaneously express IFNγ and IL-17 (21).

In the context of these observations, and given that *IL23R* and *IL12RB2* are contiguous genes sharing regulatory regions, it is likely that polymorphic variants at this genetic locus contribute to the regulation of coordinated gene expression affecting the Th1:Th17

ratio and to the synergistic or antagonistic interactions between these cell types in patients with BD. *IL23R* generates at least 6 alternatively spliced messenger RNA (mRNA), and *IL12RB2* generates at least 2 alternatively spliced mRNA, leading to diverse isoforms of the receptor proteins (23,24). Although the SNPs found associated with BD in the *IL23R–IL12RB2* locus are synonymous variants or located in noncoding regions, a variety of mechanisms may underlie susceptibility to the disease within these loci, including effects on gene expression through differences in transcription factor or microRNA binding or in mRNA splicing, transport, or stability, and effects on protein folding and stability (25,26). Even subtle changes in *IL23R* and/or *IL12RB2* mRNA levels and transcripts may drive a different activation and differentiation of immune cell types in varying environments, resulting in the development of different diseases.

Interestingly, numerous polymorphisms in *IL23R* have been associated with several autoimmune diseases (27). The *IL23R* R381Q variant has been shown to exert its protective effect against immune-mediated diseases through selective reduction of IL-23–induced Th17 cell effector function (as measured by IL-17A production), without affecting Th17 cell differentiation (28). SNP rs7517847 has been associated with Crohn's disease and inflammatory bowel disease (29–31), but, to the best of our knowledge, rs4655679, rs17375018, and rs924080 have not been associated with other autoimmune disorders or other diseases (National Human Genome Research Institute catalog of GWAS; http://www.nhgri.nih.gov), and their effect may therefore be specific to BD. It would be interesting to assess whether there is any correlation between genotype status at these genetic markers and expression of *IL23R* isoforms and IL-17A levels in Th17 cells.

Alternatively, the causal polymorphism may not, as yet, have been directly studied. Given that we genotyped a relatively high number of tag SNPs in the *IL23R–IL12RB2* region, we imputed numerous neighboring polymorphisms with high confidence and identified 17 additional SNPs strongly associated with BD. Genome-wide imputation is currently performed routinely in GWAS to increase power, but "local" imputation may also be a powerful approach to detect novel associations when a dense map of observed genotypes is available. We validated this local imputation approach in our data set through direct genotyping.

Continued stimulation of activated murine Th17 cells with transforming growth factor β and IL-6 results in the production of IL-10, a pleiotropic and potent antiinflammatory cytokine (32). The IL-10–driven containment and suppression of inflammatory responses may constitute a mechanism for autoregulation of Th17 activity upon prolonged activity. Elevated IL-10 levels have been observed in the serum and mucocutaneous lesions of patients with BD (18,33). In the 2 recently published GWAS analyses of patients with BD (1,2), 5 SNPs in *IL10* were reported to be associated with BD: rs1554286, rs1518111, rs3024490, rs1800872, and rs1800871. Both in Turkish subjects (1) and in Japanese subjects (2), these 5 polymorphisms are clearly associated with BD ($P \leq 1.0 \times 10^{-6}$), but the strong LD between these SNPs makes it impossible to refine the association to the causal SNP.

Among these 5 SNPs, our haplotype-tagging approach led us to investigate rs1518111 and rs1554286 ($r^2 = 0.70$, CEU HapMap sample); these were chosen only because the other 3 SNPs were in strong LD with them ($r^2 \geq 0.93$). We observed a trend toward association with BD at these 2 markers, and there was 89–100% power to detect these associations in our data set. The meta-analysis of rs1518111 showed some heterogeneity between studies, due to differences in the MAF values in each of the populations with different ethnic backgrounds. The results of this replication study have shown the value of exploring the genetic diversity in different ethnic groups to validate and refine genetic associations. A search for additional variants by deep-sequencing the *IL10* locus may constitute an interesting avenue for following up these findings.

Our association findings are not likely to be confounded by hidden population stratification, since they were adjusted for a very subtle population substructure. It is noteworthy that the adjusted and unadjusted association results were very similar. Even though the panel of AIMs used may not have enough power to detect patterns of variation within the Iranian genetic pool, it could be used to statistically determine whether cases and controls are matched for European, Asian, and African ancestries, and to assess whether the self-reported ancestries are correlated with any underlying genetic population substructure.

Variants in *IL10* (rs3024505 and rs3024493) have been associated with ulcerative colitis and systemic lupus erythematosus (34,35). These 2 variants are in complete LD in the CEU HapMap sample ($r^2 = 1.00$), and rs3024505 (SNP 1) demonstrated no association with BD in our data set or in the Turkish population (1). Several SNPs in *IL23R* have been associated with inflammatory bowel disease, Crohn's disease, psoriasis, psoriatic arthritis, ulcerative colitis, and ankylosing spondylitis (31, 36–38), and in our study, some of these polymorphisms

were associated with BD (rs7517847 and rs1343151), whereas others were not associated (rs10889677). Given the conservative nature of the immune system and the tight control of the different cell populations, it is not surprising that risk alleles of genes that are important in the immune response may constitute bona fide susceptibility genes for a number of different autoimmune conditions. The next challenge will be to design functional studies that allow us to investigate how different variants in *IL23R* and *IL10* translate into physiologic processes that influence immune disease risk.

For *IL23R*, we believe research efforts should concentrate on upstream and downstream regulatory regions, as was highlighted by our findings in the imputation analysis. Given that we directly analyzed only a small fraction of the existing variation in these genes and that the imputation analysis focused on common variants, we cannot exclude the possibility that other common variants that are in LD or rare variants have a stronger effect on the risk of BD. Therefore, a more in-depth study (e.g., deep-sequencing, functional studies) of these genes should be pursued.

Furthermore, as in any complex disease, gene–gene interactions are thought to occur and to increase risk. Even though we did not detect any epistasis between *IL10* and *IL23R–IL12RB2*, it will be of great interest to test, in a larger data set, the interactions between MHC-related genes (e.g., *HLA–B51*, *HLA–A26*), non-MHC genes (e.g., *IL10*, *IL23R–IL12RB2* locus), and non-nuclear variants (e.g., mitochondrial genome) (39). The definitive validation of these associations is of foremost importance, as it may have therapeutic implications. In particular, it may lead to the development of a therapy that specifically targets the IL-23/IL-17 axis, involving the use of monoclonal antibodies against proteins such as IL-23p19.

ACKNOWLEDGMENTS

We thank Dr. Majid Zeidi (Iranian Blood Transfusion Organization) for his valuable support. We are also thankful to Dr. Sirous Zeinali (Pasteur Institute of Iran) and Dr. Kayvan Saeedfar for their valuable help. We are also deeply grateful to all study participants, to the genotyping unit at the Instituto Gulbenkian de Ciência, and to Dr. Luís Graça for critical review of the manuscript.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Oliveira had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Xavier, Shahram, Davatchi, Rosa, Crespo, Abdollahi, Nadji, Jesus, Barcelos, Patto, Shafiee, Ghaderibarim, Oliveira.

Acquisition of data. Xavier, Shahram, Davatchi, Abdollahi, Nadji, Shafiee, Ghaderibarim, Oliveira.

Analysis and interpretation of data. Xavier, Shahram, Davatchi, Rosa, Crespo, Abdollahi, Nadji, Jesus, Barcelos, Patto, Shafiee, Ghaderibarim, Oliveira.

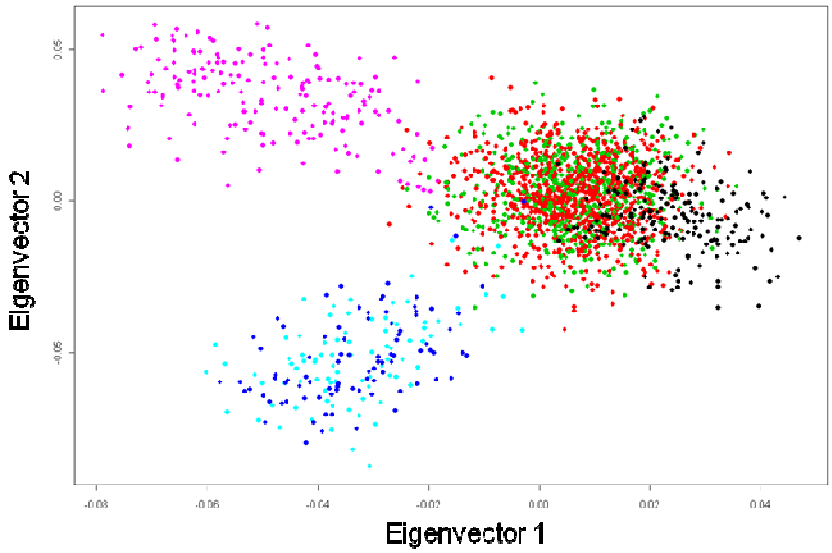
REFERENCES

1. Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet 2010;42:698–702.
2. Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, et al. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat Genet 2010;42: 703–6.
3. International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD). Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 2006; 24 Suppl 42:S14–5.
4. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–5.
5. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, et al. Discerning the ancestry of European Americans in genetic association studies. PLoS Genet 2008;4:e236.
6. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190.
7. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
8. Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, Estivill X, et al. SNPassoc: an R package to perform whole genome association studies. Bioinformatics 2007;23:644–5.
9. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209–13.
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75.
11. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 2001;69:138–47.
12. Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, et al. Behçet's disease: from East to West. Clin Rheumatol 2010;29:823–33.
13. Jiang Z, Yang P, Hou S, Du L, Xie L, Zhou H, et al. IL-23R gene confers susceptibility to Behçet's disease in a Chinese Han population. Ann Rheum Dis 2010;69:1325–8.
14. Parham C, Chirica M, Timans J, Vaisberg E, Travis M, Cheung J, et al. A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rβ1 and a novel cytokine receptor subunit, IL-23R. J Immunol 2002;168:5699–708.
15. Kastelein RA, Hunter CA, Cua DJ. Discovery and biology of IL-23 and IL-27: related but functionally distinct regulators of inflammation. Annu Rev Immunol 2007;25:221–42.
16. Korn T, Bettelli E, Oukka M, Kuchroo VK. IL-17 and Th17 cells. Annu Rev Immunol 2009;27:485–517.
17. Frassanito MA, Dammacco R, Cafforio P, Dammacco F. Th1

polarization of the immune response in Behçet's disease: a putative pathogenetic role of interleukin-12. Arthritis Rheum 1999;42: 1967–74.
18. Ben Ahmed M, Houman H, Miled M, Dellagi K, Louzir H. Involvement of chemokines and Th1 cytokines in the pathogenesis of mucocutaneous lesions of Behçet's disease. Arthritis Rheum 2004;50:2291–5.
19. Imamura Y, Kurokawa MS, Yoshikawa H, Nara K, Takada E, Masuda C, et al. Involvement of Th1 cells and heat shock protein 60 in the pathogenesis of intestinal Behçet's disease. Clin Exp Immunol 2005;139:371–8.
20. Chi W, Zhu X, Yang P, Liu X, Lin X, Zhou H, et al. Upregulated IL-23 and IL-17 in Behçet patients with active uveitis. Invest Ophthalmol Vis Sci 2008;49:3058–64.
21. Kim J, Park J, Lee E, Lee Y, Song Y, Lee E. Imbalance of Th17 to Th1 cells in Behçet's disease. Clin Exp Rheumatol 2010;28: S27–30.
22. Locksley RM. Nine lives: plasticity among T helper cell subsets. J Exp Med 2009;206:1643–6.
23. Zhang XY, Zhang HJ, Zhang Y, Fu YJ, He J, Zhu LP, et al. Identification and expression analysis of alternatively spliced isoforms of human interleukin-23 receptor gene in normal lymphoid cells and selected tumor cells. Immunogenetics 2006;57:934–43.
24. Van Rietschoten JG, Smits HH, Westland R, Verweij CL, den Hartog MT, Wierenga EA. Genomic organization of the human interleukin-12 receptor β2-chain gene. Immunogenetics 2000;51: 30–6.
25. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 2007;315:525–8.
26. Bartoszewski RA, Jablonsky M, Bartoszewska S, Stevenson L, Dai Q, Kappes J, et al. A synonymous single nucleotide polymorphism in ΔF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. J Biol Chem 2010;285: 28741–8.
27. Safrany E, Melegh B. Functional variants of the interleukin-23 receptor gene in non-gastrointestinal autoimmune diseases. Curr Med Chem 2009;16:3766–74.
28. Di Meglio P, Di Cesare A, Laggner U, Chu CC, Napolitano L, Villanova F, et al. The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans. PLoS One 2011;6:e17160.
29. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007;39:596–604.
30. Amre DK, Mack D, Israel D, Morgan K, Lambrette P, Law L, et al. Association between genetic variants in the IL-23R gene and early-onset Crohn's disease: results from a case-control and family-based study among Canadian children. Am J Gastroenterol 2008;103:615–20.
31. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461–3.
32. McGeachy MJ, Bak-Jensen KS, Chen Y, Tato CM, Blumenschein W, McClanahan T, et al. TGF-β and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain T_H-17 cell-mediated pathology. Nat Immunol 2007;8:1390–7.
33. Turan B, Gallati H, Erdi H, Gurler A, Michel BA, Villiger PM. Systemic levels of the T cell regulatory cytokines IL-10 and IL-12 in Behçet's disease: soluble TNFR-75 as a biological marker of disease activity. J Rheumatol 1997;24:128–32.
34. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008;40:1319–23.
35. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet 2009;41:1228–33.
36. Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proc Natl Acad Sci U S A 2007;104:14747–52.
37. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009; 41:216–20.
38. Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, Pointon JJ, et al, for the Australo-Anglo-American Spondyloarthritis Consortium (TASC). Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 2010;42:123–7.
39. Xavier JM, Shafiee NM, Ghaderi F, Rosa A, Abdollahi BS, Nadji A, et al. Association of mitochondrial polymorphism m.709G>A with Behçet's disease. Ann Rheum Dis 2011;70:1514–6.

4.2 SUPPLEMENTARY MATERIAL

A.



B.

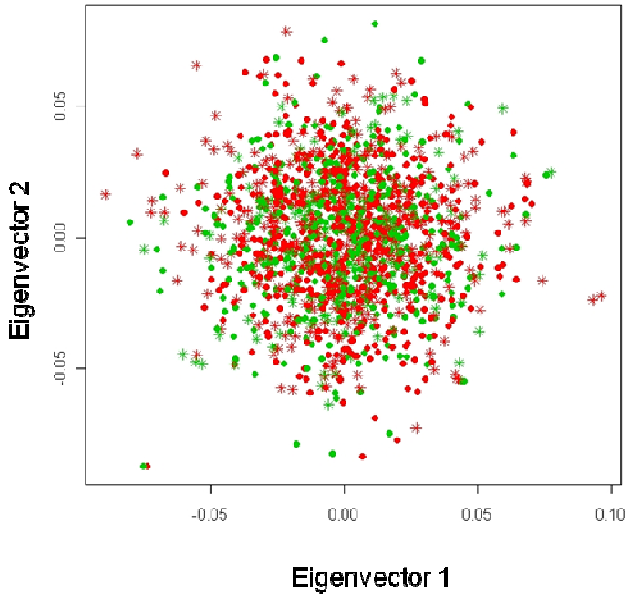


Figure 1 Top two axes of variation of the 973 Iranian BD patients and 637 controls with (A) and without (B) HapMap reference samples(174 CEU, 86 CHB, 89 JPT and 176 YRI individuals), using the genotypes at 86 ancestry informative markers. BD cases are shown in red, BD controls in green, CEU in black, YRI in pink, CHB in light blue, and JPT in dark blue. In the plot with the Iranian samples only, the discs and stars represent individuals of self-reported West-Eurasian Caucasoid and East/Central-Asian ancestries, respectively. This figure has been shown previously in the published paper (chapter 4.1, page 78) but given its poor printing quality it is shown again in this section with better quality.

Supplementary Table 1. Ancestry informative markers (AIM) used in the population stratification analysis. The chromosome (Chr.) and position in dbSNP131 are shown for all SNPs.

AIM ID	SNP ID	Informativeness rank*	Chr.	Position (bp)	Comments
1	rs11807062	91	1	3153237	Genotype data not available in CHB and JPT HapMap samples
2	rs495347	48	1	18163922	
3	rs1416467	32	1	80758728	
4	rs7552548	99	1	83063846	
5	rs1890131	11	1	167597473	
6	rs725974	15	1	170613251	
7	rs2236876	47	1	173017486	
8	rs2419063	88	1	190256835	
9	rs1157492	21	1	215238904	
10	rs6432110	27	2	10738018	
11	rs4832640	79	2	19104143	
12	rs1517407	31	2	63490331	
13	rs1364394	3	2	96195730	Out of HWE
14	rs6745653	51	2	98289997	
15	rs10496610	19	2	124041452	SNP not included in the multiplex
16	rs3769005	2	2	136603366	
17	rs1448314	41	2	223986200	
18	rs2596834	14	3	10820067	
19	rs822759	29	3	22973112	
20	rs1879558	56	3	153654355	
21	rs9290675	26	3	179006633	
22	rs4859259	97	3	182664547	
23	rs4686497	44	3	188641507	
24	rs9861816	85	3	194168500	Genotype data not available in CHB and JPT HapMap samples
25	rs9328764	96	4	2176454	
26	rs2014303	77	4	10585856	
27	rs1873195	68	4	38891173	
28	rs17443616	8	4	41222165	
29	rs4555709	83	4	54026961	
30	rs1922286	9	4	80357377	
31	rs10516982	42	4	96606101	
32	rs974020	6	4	117507699	
33	rs1073321	53	4	118512852	
34	rs12502036	95	4	158609743	Out of HWE
35	rs1373557	30	4	167244207	
36	rs2251432	66	4	190688592	
37	rs16891982	12	5	33951693	
38	rs33706	20	5	87505164	SNP not included in the multiplex
39	rs3822616	61	5	94802389	
40	rs153595	57	5	115684182	
41	rs2804756	75	6	753586	
42	rs10484547	72	6	29452774	
43	rs2596501	4	6	31321211	Out of HWE
44	rs2187684	36	6	32764719	
45	rs756147	69	6	141630719	
46	rs1032143	98	6	155486182	
47	rs2171209	7	6	159183562	
48	rs2097884	94	7	4352417	
49	rs10486207	76	7	8110780	
50	rs2905347	38	7	22620319	
51	rs17864053	89	7	90157306	

52	rs2219248	55	7	114372026	Renamed rs647055; SNP not included in the multiplex
53	rs2367191	40	7	142235861	
54	rs1922086	25	7	156016226	
55	rs920590	43	8	19651161	
56	rs4639533	90	8	53193238	
57	rs10504924	84	8	94088261	SNP not included in the multiplex
58	rs2086085	34	9	1690761	Failed genotyping quality control
59	rs1408794	58	9	12651340	
60	rs10512122	33	9	84485970	
61	rs10508372	13	10	8972018	
62	rs1045873	64	10	25137772	
63	rs7908825	80	10	75554541	
64	rs10509384	46	10	79023182	
65	rs379773	24	10	109915352	
66	rs10509954	65	10	113668388	
67	rs1560569	10	11	8939887	
68	rs923031	60	11	15821026	SNP not included in the multiplex
69	rs7108371	62	11	44552724	
70	rs4938377	100	11	117296292	SNP not included in the multiplex
71	rs2847502	71	11	120118498	
72	rs1003306	35	12	2754197	
73	rs7965049	50	12	7645777	
74	rs998401	59	12	55355233	
75	rs3809125	63	12	56844349	
76	rs1582398	52	12	80121551	
77	rs2047058	37	12	128200960	
78	rs986642	54	13	58046262	
79	rs7997100	22	13	69090074	
80	rs1854226	17	13	98238242	Failed genotyping quality control
81	rs10483853	78	14	73756299	
82	rs1777689	81	14	83125519	
83	rs1129038	1	15	28356859	Genotype data not available in all HapMap samples
84	rs7163907	74	15	75845097	Failed genotyping quality control
85	rs10519269	92	15	79918378	
86	rs8041327	23	15	86809807	
87	rs1476162	49	17	8091083	
88	rs1107820	87	17	44135359	
89	rs2003092	70	17	62162777	
90	rs959260	28	17	73369422	
91	rs523776	67	18	7564299	
92	rs2418844	16	18	27783263	
93	rs959763	39	18	58646846	Failed genotyping quality control
94	rs4892082	93	18	70900692	
95	rs10853962	86	19	2897881	
96	rs103294	18	19	54797848	
97	rs202546	45	20	1663539	
98	rs477627	82	20	48180058	
99	rs7277342	73	21	30118974	Failed genotyping quality control
100	rs969539	5	22	26617787	

*http://genepath.med.harvard.edu/~reich/EUROSNP.htm; QC: quality control.

Supplementary Table 2. FST statistics between Iranian BD cases and controls, and HapMap CEU, CHB, JPT, and YRI data sets.

	Controls	**CEU**	**CHB**	**JPT**	**YRI**
Cases	0.000	0.073	0.155	0.157	0.186
Controls		0.075	0.157	0.159	0.187
CEU			0.252	0.251	0.315
CHB				0.006	0.222
JPT					0.227

Supplementary Table 3. Identification and basic characterization of the polymorphisms investigated in this study.

Gene/locus	SNP ID	SNP	Position (bp)	Alleles 1:2	HWE	Frequency of allele 1 in cases	Frequency of allele 1 in controls
IL10	1	rs3024505	206939904	T:C	0.970	0.142	0.133
	2	rs3024498	206941529	G:A	0.033	0.138	0.136
	3	rs1554286	206944233	T:C	0.695	0.250	0.220
	4	rs1518111	206944645	A:G	0.952	0.318	0.280
	5	rs1800896	206946897	C:T	0.126	0.343	0.339
	6	rs1800890	206949365	A:T	0.134	0.266	0.259
IL23R-IL12RB2	7	rs4655679	67599657	C:T	0.889	0.804	0.738
	8	rs9729046	67607082	C:G	0.880	0.713	0.662
	9	rs6683039	67631333	T:C	0.806	0.561	0.551
	10	rs17375018	67655147	G:A	0.841	0.797	0.723
	11	rs6656929	67661041	A:T	1.000	0.683	0.628
	12	rs10489630	67662622	T:G	0.736	0.672	0.617
	13	rs11209018	67667291	A:G	0.954	0.531	0.506
	14	rs7539625	67672765	A:G	0.567	0.481	0.441
	15	rs790631	67676922	C:T	0.272	0.269	0.228
	16	rs790633	67678993	T:C	0.222	0.289	0.250
	17	rs7517847	67681669	T:G	0.023	0.755	0.675
	18	rs7530511	67685387	T:C	0.728	0.128	0.090
	19	rs10489629	67688349	A:G	0.644	0.674	0.614
	20	rs4655692	67691665	A:G	0.763	0.205	0.170
	21	rs12030948	67701765	T:G	0.116	0.499	0.498
	22	rs10489628	67704107	C:T	0.757	0.702	0.666
	23	rs10789229	67705574	T:C	0.140	0.700	0.695
	24	rs6682033	67708670	A:G	0.941	0.822	0.782
	25	rs1343151	67719129	C:T	0.595	0.788	0.729
	26	rs6693831	67720867	T:C	0.414	0.243	0.203
	27	rs11465817	67721097	A:C	0.487	0.410	0.390
	28	rs10889675	67722216	A:C	0.727	0.118	0.113
	29	rs10889677	67725120	A:C	0.107	0.467	0.444
	30	rs11209030	67737775	C:A	0.664	0.828	0.787
	31	rs17303361	67739824	A:G	0.518	0.898	0.885
	32	rs1495965	67753508	G:A	0.657	0.587	0.516
	33	rs924080	67760140	T:C	0.415	0.687	0.612
	34	rs7539817	67767993	G:A	0.066	0.653	0.651
	35	rs10889680	67768593	T:A	0.668	0.145	0.137
	36	rs12131065	67769006	G:A	0.198	0.755	0.754
	37	rs11209045	67770799	A:G	0.126	0.655	0.645
	38	rs3790558	67777021	T:G	0.079	0.569	0.568
	39	rs10489627	67779676	C:T	0.223	0.316	0.314
	40	rs1890741	67782752	C:G	0.448	0.168	0.167
	41	rs2201584	67787715	C:T	0.615	0.867	0.865
	42	rs6693065	67800018	A:G	0.060	0.784	0.779
	43	rs1995147	67803059	G:C	0.200	0.412	0.410
	44	rs3790565	67811356	T:C	0.954	0.796	0.796
	45	rs6679356	67820194	C:T	0.989	0.189	0.179
	46	rs3790567	67822377	A:G	0.687	0.280	0.267
	47	rs3790569	67836860	T:C	0.049	0.837	0.835
	48	rs2229546	67861520	A:C	0.485	0.683	0.680

CHAPTER 5

Gene expression profiling and association studies implicate the neuregulin signalling pathway in Behçet's disease susceptibility

(Xavier *et al.,* J Mol Med (Berl). 2013)

5.1 MAIN MANUSCRIPT

J Mol Med
DOI 10.1007/s00109-013-1022-4

JMolMed

ORIGINAL ARTICLE

Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility

Joana M. Xavier · Tiago Krug · Fereydoun Davatchi · Farhad Shahram · Benedita V. Fonseca · Gorete Jesus · Filipe Barcelos · Joana Vedes · Manuel Salgado · Bahar Sadeghi Abdollahi · Abdolhadi Nadji · Maria Francisca Moraes-Fontes · Niloofar Mojarad Shafiee · Fahmida Ghaderibarmi · José Vaz Patto · Jorge Crespo · Sofia A. Oliveira

Received: 10 December 2012 / Revised: 1 February 2013 / Accepted: 7 March 2013

Abstract Behçet's disease (BD) is a complex disease with genetic and environmental risk factors implicated in its etiology; however, its pathophysiology is poorly understood. To decipher BD's genetic underpinnings, we combined gene expression profiling with pathway analysis and association studies. We compared the gene expression profiles in peripheral blood mononuclear cells (PBMCs) of 15 patients and 14 matched controls using Affymetrix microarrays and found that the neuregulin signaling pathway was over-represented among the differentially expressed genes. The Epiregulin (*EREG*), Amphiregulin (*AREG*), and Neuregulin-1 (*NRG1*) genes of this pathway stand out as they are also among the top differentially expressed genes. Twelve haplotype tagging SNPs at the *EREG-AREG* locus and 15 SNPs in *NRG1* found associated in at least one published BD genome-wide association study were tested for association with BD in a dataset of 976 Iranian patients and 839 controls. We found a novel association with BD for the rs6845297 SNP located downstream of *EREG,* and replicated three associations at *NRG1* (rs4489285, rs383632, and rs1462891). Multifactor dimensionality reduction analysis indicated the existence of epistatic interactions between *EREG* and *NRG1* variants. *EREG-AREG* and *NRG1*, which are members of the epidermal growth factor (EGF) family, seem to modulate BD susceptibility through main effects and gene–gene interactions. These association findings support a role for the EGF/ErbB signaling pathway in

Electronic supplementary material The online version of this article (doi:10.1007/s00109-013-1022-4) contains supplementary material, which is available to authorized users.

J. M. Xavier · T. Krug · B. V. Fonseca · S. A. Oliveira (✉)
Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Av. Prof. Egas Moniz,
Edifício Egas Moniz,
1649-028 Lisboa, Portugal
e-mail: aaoliveira@fm.ul.pt

J. M. Xavier · T. Krug · B. V. Fonseca · S. A. Oliveira
Instituto Gulbenkian de Ciência, Oeiras, Portugal

F. Davatchi · F. Shahram · B. S. Abdollahi · A. Nadji · N. M. Shafiee · F. Ghaderibarmi
Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran

G. Jesus
Hospital Infante D. Pedro, Aveiro, Portugal

F. Barcelos · J. Vaz Patto
Instituto Português de Reumatologia,
Lisboa, Portugal

J. Vedes
Hospital de Sousa Martins, Guarda, Portugal

M. Salgado
Hospital Pediátrico de Coimbra,
Coimbra, Portugal

M. F. Moraes-Fontes
Hospital Curry Cabral, Lisboa, Portugal

J. Crespo
Hospitais da Universidade de Coimbra,
Coimbra, Portugal

Published online: 27 April 2013

BD pathogenesis that warrants further investigation and highlight the importance of combining genetic and genomic approaches to dissect the genetic architecture of complex diseases.

Keywords Genetics · Microarray · Association study · Behçet's disease · Genetic epidemiology

Introduction

Behçet's disease (BD) is a systemic immuno-inflammatory disease characterized by the peculiar form it affects the organism, especially in the mucocutaneous and ocular structures. Its most common clinical manifestations are recurrent orogenital ulcerations and inflammatory eye lesions. Although its etiopathogenesis remains to be elucidated, genetic (e.g., HLA-B51) and environmental factors as well as immune-mediated mechanisms have been pointed out as important factors [1].

Genome-wide association studies (GWAS) for BD recently revealed two novel susceptibility loci at a genome-wide significant level: *IL23R-IL12RB2* and *IL10* [2, 3]. However, the genetic susceptibility to BD is governed by additional genes that cannot be pinpointed using this approach. To identify additional genetic risk factors for BD, we hereby pursued another strategy which intersects data from a genome-wide expression study with an in silico pathway analysis and association studies.

Gene expression profiling allows the simultaneous monitoring of the transcriptional behavior of thousands of genes, and genes differentially expressed between patients and controls in a relevant tissue are likely to be involved at some stage of the pathogenic process. Gene profiling experiments have allowed the identification of novel and unsuspected pathways involved in disease etiology, including in rheumatic and immune diseases [4–6]. Pathway analysis of differentially expressed genes is a useful approach to highlight undetected patterns in the data and to determine which differentially expressed genes to follow-up in independent assays such as association studies. This convergence of genomic and genetic approaches enables data mining beyond the top findings. With this strategy, we were able to highlight the role of the neuregulin signaling pathway in BD pathogenesis and strengthen the previously observed association of variants in *NRG1* with BD.

Materials and methods

Gene profiling study subjects

Fourteen BD patients and 15 healthy controls were enrolled in the gene expression study through the Hospital Infante D. Pedro, Hospital de Sousa Martins, Hospital Pediátrico de Coimbra and Instituto Português de Reumatologia in Portugal. These participants are of Portuguese Caucasian origin. All patients satisfied the International Criteria for Behçet's disease (ICBD) [7]. Patients were considered immunosuppressed if medicated with azathioprine, cyclosporine, deflazacort, or thalidomide. BD patients with age-at-diagnosis after 60 years were excluded. Controls were evaluated using the same evaluation procedures as the cases and selected when negative for BD and any other rheumatologic or autoimmune disorder. This research was approved by the ethics committee at Hospital Infante D. Pedro and Instituto Português de Reumatologia where the samples were collected. All participants were informed of the study and provided informed written consent.

Total RNA isolation and microarray hybridization

Whole blood samples were obtained by venipuncture and collected in BD Vacutainer CPT tubes (BD, Franklin Lakes, NJ, USA). These samples were centrifuged to isolate PBMCs, which were then washed twice. Total RNA was extracted from PBMCs using RNeasy Mini kit (Qiagen, Hilden, Germany) according to the protocol recommended by the manufacturer. Total RNA (3.5 μg) from each individual was hybridized to GeneChip Human Genome U133 Plus 2.0 microarrays (Affymetrix, Santa Clara, CA, USA) at the Instituto Gulbenkian de Ciência's Affymetrix Core Facility following manufacturer's protocols. Extensive quality control checks were performed in all steps of the process.

Gene expression data normalization and statistical analysis

The intensity array data were analyzed with their respective CDF file from Affymetrix on the Partek software (Partek Incorporated, St Louis, MO, USA). Background correction, normalization, and summarization of the CEL files were performed using the robust multichip average algorithm. Differentially expressed genes among cases and controls were identified with analysis of variance (ANOVA), adjusting for known experimental (immunosuppression status) and study design (geographic origin and scan date) covariates. The genes with more than 1.20-fold-change and $P \leq 0.05$ were considered differentially expressed. To account for multiple testing, we calculated false discovery rates (FDR) using the Partek Q value method [8].

The gene expression profiling was conducted and reported in accordance with the minimum information about microarray experiment (MIAME) guidelines [9]. The microarray data were deposited on Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/projects/geo/) with the accession number GSE17114.

Hierarchical clustering (HC) analysis was performed using the Partek software. The effects of the geographic origin of the participants, the scan date of the microarrays

and the immunosuppression status of the participants were removed using the batch-remove tool implemented in Partek prior to visualization, since these visualization tools cannot correct for study design batch effects.

Quantitative real-time polymerase chain reaction confirmation of some of the microarray results was not performed because of an insufficient amount of RNA available for several of the controls and patients used in this study.

Pathway analyses

Ingenuity Pathway Analysis 9.0 (IPA; Ingenuity Systems, http://www.ingenuity.com) was used to functionally annotate genes according to molecular networks, biological function, and canonical pathways. An IPA list is built upon a very large manually curated and up-to-date database of genes, proteins, functions, interactions/networks, and pathways. Functional analysis was performed on the 621 probe IDs differentially expressed ($P \leq 0.05$ and absolute fold-change>1.20). The IPA "Core Analysis" was run using the Human Genome U133 Plus 2.0 array as the reference set, using direct and indirect relationships for network analysis, including endogenous chemicals, data from Human species only, experimental observed confidence, and all tissues, cell lines, and data sources. The significance (P values) of the association between a dataset and a canonical pathway was determined by comparing the number of genes in a dataset that participate in a given pathway to the total number of occurrences of these genes in all pathway annotations. A Fisher's exact test was used to calculate the P value to determine the probability that the association between the genes in the dataset and the canonical pathway is explained only by chance. The level of statistical significance was set to $P \leq 0.05$.

Association study subjects

Patients (976) and 839 controls, ascertained and collected as described previously [10], were included in the association study. Diagnosis of BD was made according to the ICBD criteria [7]. This research was approved by the ethics committee at the Tehran University for Medical Sciences, Iran. All participants were informed of the study and provided informed written consent.

Genotyping

Thirteen haplotype tagging SNPs (htSNP) at the *EREG-AREG* locus (chr. 4: 75449724–75545341 bp) were identified in Haploview 4.1 [11] using genotypes of 30 European (CEU) trios downloaded from the HapMap Release 24/phase II Nov08 (NCBI B36 assembly, dbSNP b126), and with the following options: pairwise mode, $r^2>0.80$, and minor allele frequency (MAF)>0.1. Fifteen SNPs with a $r^2<0.8$ in *NRG1* (chr. 8: 31517663–32066776 bp) found associated with BD in either one of two published GWAS (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki) were genotyped.

Genomic DNA was extracted from whole blood samples using a salting out procedure. SNPs were genotyped using Sequenom's (San Diego, USA) iPlex assay (primer extension of multiplex products with detection by matrix assisted laser desorption/ionization time-of-flight mass spectrometry) following the manufacturer's protocol and detected in a Sequenom MassArray K2 platform. The primer sequences are available in Supplementary Table 1 and were designed using Sequenom's MassARRAY ® Assay Design 3.0 software. The genotyping was performed at the Instituto Gulbenkian de Ciência's Genomics Unit.

Extensive quality control was performed using eight HapMap (http://hapmap.ncbi.nlm.nih.gov/) controls of diverse ethnic affiliation, sample duplication within and across plates, Mendelian inheritance check in three large pedigrees, Hardy–Weinberg equilibrium (HWE) in the control group ($P>0.01$), and a minimum of 95 % call rate for each SNP. Genotype determinations were performed blinded to affection status.

Association analyses

Student's unpaired t tests and chi-square tests were used to compare quantitative and qualitative clinical and demographic data, respectively, between BD cases and controls. χ^2 tests for HWE in controls and crude allelic association of SNPs with BD risk were performed using Haploview 4.1. The SNPassoc v.1.4-9 package [12] implemented in the R freeware (http://cran.r-project.org/) was used to assess the associations with BD adjusted for gender. Odds ratios (ORs) and their associated 95 % confidence intervals (CIs) were calculated for SNPs with significant allelic associations with BD. Results were considered significant below the conventional level of 0.05.

Meta-analyses

GWAS data for *EREG-AREG* and *NRG1* SNPs (e.g., allele counts, OR, CI) was obtained by personal communications from Drs. Ahmet Gul and Elaine F. Remmers for the Turkish dataset [2], and from Drs. Shigeaki Ohno, Nobuhisa Mizuki and Akira Meguro for the Japanese dataset [3]. Fixed effects (Mantel-Haenszel) meta-analyses were performed using the rmeta package in R.2.7.2 and PLINK v1.047.

Gene–gene interactions

Epistasis between *EREG-AREG* and *NRG1* variants was detected and characterized using the multifactor dimensionality reduction (MDR) method (v2.0 beta 8.3) (www.epistasis.org) [13]. Missing genotypes were first imputed for each SNP using

the PLINK database (http://pngu.mgh.harvard.edu/purcell/plink/) [14] with genotypes from individuals in the CEU HapMap population as a reference. MDR reduces high-dimensional data into a single dimension data by pooling multilocus genotypes into "high- risk" and "low-risk" groups according to ratio of affected and unaffected individuals within each genotype combination. This new multilocus variable is then tested for its ability to predict disease status. Cross validation and testing balance accuracy (TBA) are used to select the best models [13, 15]. The statistical significances of the best models were calculated after 1,000 permutations using the MDR Permutation Tool (v1.0, beta 2). Entropy-based interaction dendrograms were drawn to interpret epistasis models.

Results

Gene expression study

The principal clinical and demographic characteristics of the 29 individuals used in the gene profiling study are shown in Table 1. Cases and controls were matched for age-at-examination (AEE) (mean AAE±SD of 37.1±11.0 years in cases and 36.7±13.5 years in controls, P=9.39E-01). All of the microarrays performed were of good quality (average±SD of present calls and of background were 45.0±3.0 % and 42.0±6.1 %, respectively).

Using analysis of variance on the normalized expression data, 621 probe sets representing 508 genes (Supplementary Table 2) were found differentially expressed among BD cases and controls with a threshold of 1.20-fold-change, and $P\leq$0.05. 373 of these probe sets (representing 314 genes) were down-regulated. All the probe sets with $P\leq$0.05 also had a Q value≤0.05. The false discovery rate was determined based on Q values, as these have a higher apparent power when compared to other standard methods [8]. The hierarchical cluster diagram shows a distinct gene expression profile between patients and controls that clearly separates the two groups of samples (Fig. 1). The top 30 genes with larger fold-change differences between cases and controls are shown in Table 2. *EREG*, *S100B* (S100 calcium binding protein B), *NAMPT* (nicotinamide phosphoribosyltransferase), *AREG,* and *NRG1* were the top five genes under-expressed in patients, and *HBG1/HBG2* (hemoglobin subunit gamma-1/gamma-2), *SRSF6* (serine/arginine-rich splicing factor 6), *LGALS2* (lectin, galactoside-binding, soluble, 2), *NEBL* (nebulette), and *TRD@* (T cell receptor delta locus) the top five over-expressed genes in BD patients.

Pathway analysis

Pathway analysis of the differentially expressed genes was performed using the Ingenuity Pathway Analysis (IPA). This software enables the discovery and analyses of functional relationships between differentially expressed genes and gives information about the molecular networks, biological functions, and canonical pathways over-represented among those genes.

Fourteen canonical pathways (Supplementary Table 3) were found to be significantly over-represented among the differentially expressed genes, including the lipid antigen presentation by CD1 (4/22 genes, P=4.34E-04) and antigen presentation (4/40 genes, P=9.12E-03) pathways which have functions related to cell-mediated immunity. IPA analysis also highlighted several pathways related to cell signaling such as the RhoA signaling (9/22 genes, P=1.72E-03), Cdc42 signaling (9/143 genes, P=3.51E-03), JAK2 in hormone-like cytokine signaling (4/34 genes, P=8.17E-03), caveolar-mediated endocytosis signaling (6/81 genes, P=8.88E-03), and neuregulin signaling (6/95 genes, P=2.04E-02). Interestingly, three genes of the neuregulin signaling pathway—*EREG* (−2.39 fold-change, P=2.99E-02), *AREG* (−1.80-fold-change, P=4.90E-02), and *NRG1* (−1.76-fold-change, P=3.12E-02)—were among the most differentially expressed genes between cases and controls, and therefore these three genes were selected for follow up in an association study.

Association study

To evaluate the association between genes of the neuregulin signaling pathway (*EREG-AREG* and *NRG1*) and susceptibility to BD, we performed a case–control association study. The general characteristics of the association study dataset are shown in Table 3. The distribution of clinical symptoms in the Iranian BD patients (e.g., 98.7 % with oral aphthosis, 62.3 % with genital aphthosis, 55.2 % with skin lesions, 59.7 % with ocular lesions) is in line with what has been observed for larger datasets [16], suggesting that this is a representative group. Cases and controls were matched for age-at-examination (mean AAE±SD of 39.1±11.0 years in BD and 40.0±12.3 years in controls, P=1.93E-01).

Since no SNPs in the *EREG-AREG* locus showed evidence of association in the two published BD GWAS (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki), we followed a haplotype tagging SNP approach in order to explore the association of this locus with BD. In an extended region of chromosome 8 (chr. 8: 31517663–32713793 bp), including *NRG1,* 17 and 22 SNPs have been associated with BD in the Japanese and Turkish populations (P<0.05), respectively (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki). Since *NRG1* is a very large gene spanning over 1.1 Mb, we focused on 15 independent SNPs located within the smaller region showing stronger evidence of association (chr. 8: 31517663–32066776 bp). In the *EREG-AREG* region, one

Table 1 Characterization of the sample used in the gene profiling study

ID	Affection status	Sex	AAE (years)	AAD (years)	Major clinical symptoms	IS
1	Case	F	29	21	Oral and genital aphtosis, pseudofolliculitis	Yes
2	Case	F	40	35	Oral and genital aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis	Yes
3	Case	F	36	35	Oral and genital aphtosis, pseudofolliculitis, positive pathergy	No
4	Case	F	29	21	Oral and genital aphtosis	No
5	Case	F	55	32	Oral and genital aphtosis, erythema nodosum, positive pathergy	No
6	Case	F	30	29	Oral and genital aphtosis, erythema nodosum, positive pathergy	Yes
7	Case	F	44	30	Oral and genital aphtosis, pseudofolliculitis and posterior uveitis	Yes
8	Case	F	46	44	Oral and genital aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis	No
9	Case	M	20	17	Oral and genital aphtosis, pseudofolliculitis	Yes
10	Case	M	57	36	Oral aphtosis, pseudofolliculitis, erythema nodosum, posterior uveitis	No
11	Case	M	50	30	Oral and genital aphtosis, erythema nodosum, large vein thrombosis	Yes
12	Case	M	30	30	Oral aphtosis, pseudofolliculitis, erythema nodosum, large vein thrombosis, positive pathergy	Yes
13	Case	M	33	21	Oral, genital and skin aphtosis, pseudofolliculitis, erythema nodosum, anterior and posterior uveitis	Yes
14	Case	M	29	27	Oral and genital aphtosis, pseudofolliculitis	No
15	Case	M	28	24	Oral and genital aphtosis, erythema nodosum, anterior and posterior uveitis, retinal vasculitis	Yes
		Mean±SD	37.1±11.0	28.8±7.2		
16	Control	F	27	–	–	No
17	Control	F	32	–	–	No
18	Control	F	62	–	–	No
19	Control	F	51	–	–	No
20	Control	F	26	–	–	No
21	Control	F	26	–	–	No
22	Control	F	46	–	–	No
23	Control	M	35	–	–	No
24	Control	M	42	–	–	No
25	Control	M	31	–	–	No
26	Control	M	28	–	–	No
27	Control	M	61	–	–	No
28	Control	M	26	–	–	No
29	Control	M	21	–	–	No
		Mean±SD	36.7±13.5	–		

The principal demographic and clinical characteristics of the study subjects are indicated, including the age-at-examination (AAE) for the patients and controls and the age-at-diagnosis (AAD) for the patients. Patients were considered immunosuppressed (IS) if medicated with azathioprine, cyclosporine, deflazacort, or thalidomide at the time of blood sample collection

SNP failed genotyping (rs2132065) and was not included in the analysis. All the remaining SNPs were successfully genotyped (Supplementary Table 4), and the association results are shown in Supplementary Table 5.

In the *EREG-AREG* locus, we found an allelic association at rs6845297 (P=2.51E-02, OR_A [95%CI]=0.86 [0.75–0.98]), located downstream of *EREG*. In *NRG1*, three markers demonstrated an association with BD risk in the allelic test: rs4489285 (P=2.01E-02, OR_G [95%CI]=0.85 [0.73–0.97]) located upstream *NRG1*, and rs383632 (P=2.78E-02, OR_T [95%CI]=1.22 [1.02–1.46]) and rs1462891 (P=3.92E-02, OR_T [95 % CI]=1.18 [1.01–1.37]) located in *NRG1*. rs4489285 was also associated with BD in a Japanese sample, and rs383632 and rs1462891 were associated with BD in a Turkish dataset (personal communications by Dr. Elaine Remmers and Dr. Nobuhisa Mizuki). Since male to female

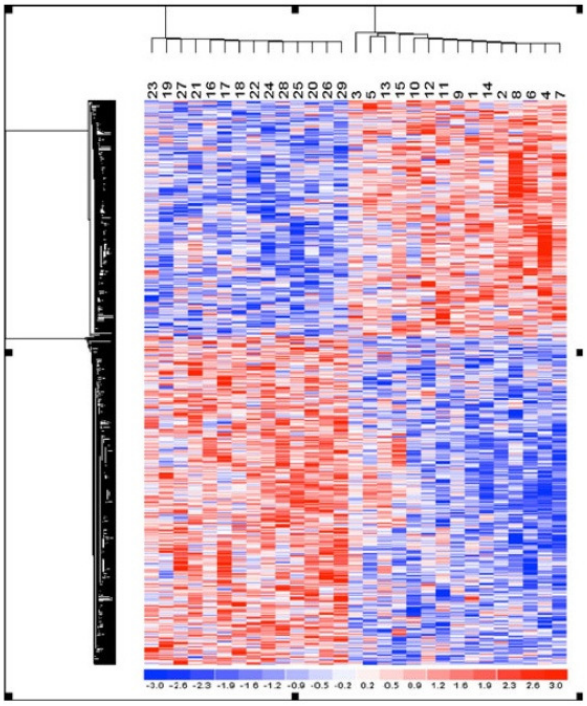


Fig. 1 Illustration of the expression pattern differences among BD cases and controls. Hierarchical clustering analysis of the 29 samples using the 621 probe sets differentially expressed between BD cases and controls, with a threshold of a 1.2-fold-change and $P \leq 0.05$. Each column represents an individual, and each row a probe set. Higher expression levels are *dark red* and lower levels are *dark blue*. Refer to Table 1 for sample identification. This figure is produced with the Partek software

ratio was significantly higher in the BD group than in the control group (52.5 % and 41.5 %, respectively, P=3.76E-06), we tested the association of these four SNPs adjusted for gender. They remained significantly associated ($P \leq 0.05$), albeit less so due to the concurrent loss of power associated with co-variate adjustment. None of these SNPs survive Bonferroni multiple testing correction. Haplotype associations did not provide additional insights over single marker associations (data not shown).

Meta-analysis of associated SNPs in *NRG1*

A combined analysis of the three associated SNPs at *NRG1* with data from two published GWAS [2, 3] reveals that the G allele of rs4489285 confers protection to BD (P=2.10E-03, OR_G [95 % CI]=0.83 [0.74–0.93]) and that the T alleles of rs383632 (P=6.61E-04, OR_T [95 % CI]=1.23 [1.09–1.38]) and rs1462891 (P=7.88E-04, OR_T[95 % CI]=1.19[1.07–1.32]) confer risk to BD (Table 4). The association consistency of these *NRG1* polymorphisms is reinforced by this meta-analysis and further supports their role in BD susceptibility.

Epistatic interaction

Given that BD has clearly a multigenic inheritance and that *EREG*, *AREG,* and *NRG1* belong to the same cellular pathway, we tested for the existence of non-additive gene–gene interactions using the multifactor-dimensionality reduction method. Supplementary Table 6 summarizes the best interaction models among all possible 2-, 3-, and 4-marker models tested. The best 4-marker interaction model was a significant predictor of Behçet's disease. The interaction between rs6845297 and rs9992496 (both located downstream of *EREG*) and rs2345991 and rs956203 (located in *NRG1*) had a statistically significant testing balanced accuracy of 0.566 (thus correctly classifying 56.6 % of the individuals tested), and a cross-validation consistency of 5/10, indicating that the model was selected five times out of ten cross-validation subsets (P=8.00E-03 after 1,000 permutations). The interaction dendogram depicted in Fig. 2 shows a strong non-linear (epistatic) synergistic interaction between rs2345991 and rs956203 (located in *NRG1*) with rs9992496 (located downstream of *EREG*). The dot-dash line indicates an independent effect of rs6845297 (located downstream of *EREG*) over the three other SNPs. These results suggest that BD susceptibility can be modulated by epistatic interactions between epidermal growth factor receptor genes.

Discussion

Using a genomic and genetic convergence strategy for the first time in the BD genetics field, the neuregulin signaling pathway emerged as a novel player. Several genes in this pathway (*EREG*, *AREG*, *NRG1*) were under-expressed in BD cases when compared to controls, and genetic markers in these loci were associated with BD susceptibility.

To the best of our knowledge, we performed the first genome-wide study comparing the expression profiles of BD patients and controls. PBMCs were considered a relevant tissue for expression profiling since BD is a generalized vasculitis characterized by a profound inflammatory and immune dysregulation [1]. Furthermore, given that BD is a complex disease with multiple genes and environmental factors implicated in its etiology, we did not expect to find genes with major expression differences. Instead, the cumulative effect of small changes in expression levels of a large number of genes, over time, will result in the phenotype. Supporting this notion is the finding that *HLA-DQA1*, a well-established locus for BD, demonstrated a moderate differential expression in our study (1.56-fold-change, P=3.19E-02). Therefore, we selected genes to follow up based on a conservative threshold of 1.20-fold-change.

Table 2 Top genes differentially expressed between BD cases and controls

Probe set ID	Gene Symbol	Gene name	*P*	Fold-change
204419_x_at	*HBG /HBG2*	Hemoglobin, gamma A/ hemoglobin, gamma G	1.35E-02	2.53
206108_s_at	*SRSF6*	Serine/arginine-rich splicing factor 6	3.96E-02	2.45
208450_at	*LGALS2*	Lectin, galactoside-binding, soluble, 2	1.11E-02	2.20
203961_at	*NEBL*	Nebulette	2.59E-02	2.10
234964_at	*TRD@*	T cell receptor delta locus	1.91E-02	1.81
222027_at	*NUCKS1*	Nuclear casein kinase and cyclin-dependent kinase substrate 1	3.06E-02	1.68
224009_x_at	*DHRS9*	Dehydrogenase/reductase (SDR family) member 9	4.92E-03	1.60
209642_at	*BUB1*	Budding uninhibited by benzimidazoles 1 homolog	4.54E-02	1.59
212671_s_at	*HLA-DQA1/2*	Major histocompatibility complex, class II, DQ alpha 1/alpha 2	3.19E-02	1.57
216733_s_at	*GATM*	Glycine amidinotransferase	8.85E-03	1.55
221530_s_at	*BHLHE41*	Basic helix-loop-helix family, member e41	4.08E-03	1.53
223751_x_at	*TLR10*	Toll-like receptor 10	2.58E-03	1.51
228285_at	*TDRD9*	Tudor domain containing 9	4.74E-02	−1.52
207314_x_at	*KIR3DL1/2 /LOC727787*	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail 1/2 /tail, 2-like	2.45E-02	−1.53
219359_at	*ATHL1*	ATH1, acid trehalase-like 1	2.40E-02	−1.53
204908_s_at	*BCL3*	B-cell CLL/lymphoma 3	1.24E-02	−1.53
216905_s_at	*ST14*	Suppression of tumorigenicity 14	3.77E-03	−1.55
227510_x_at	*MALAT1*	Metastasis associated lung adenocarcinoma transcript 1	1.12E-02	−1.57
204036_at	*LPAR1*	Lysophosphatidic acid receptor 1	3.45E-02	−1.57
202464_s_at	*PFKFB3*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	2.35E-03	−1.58
213986_s_at	*C19orf6*	Chromosome 19 open reading frame 6	5.46E-03	−1.62
1569599_at	*SAMSN1*	SAM domain, SH3 domain and nuclear localization signals 1	3.84E-02	−1.63
209264_s_at	*TSPAN4*	Tetraspanin 4	1.30E-02	−1.65
236244_at	*HNRNPU*	Heterogeneous nuclear ribonucleoprotein U	2.42E-02	−1.67
232530_at	*PLD1*	Phospholipase D1, phosphatidylcholine-specific	1.06E-02	−1.72
206343_s_at	*NRG1*	Neuregulin 1	3.12E-02	−1.76
205239_at	*AREG*	Amphiregulin	4.90E-02	−1.80
243296_at	*NAMPT*	Nicotinamide phosphoribosyltransferase	5.24E-04	−1.86
209686_at	*S100B*	S100 calcium binding protein B	7.92E-03	−2.25
205767_at	*EREG*	Epiregulin	2.99E-02	−2.39

The 30 genes with higher absolute value of fold-change difference between cases and controls are shown in this table. Genes with more than one probe differentially expressed are represented by the probe with the higher fold-change difference between cases and controls

Pathway analysis of differentially expressed genes drew our attention to the neuregulin signaling pathway and to some of its constituents, in particular to *EREG*, *AREG*, and *NRG1*, which were among the most under-expressed genes in BD patients. Association studies of these genes revealed one polymorphism in the *EREG-AREG* locus (rs6845297) and three SNPs in *NRG1* (rs4489285, rs383632, and rs1462891) associated with BD. While the association of rs6845297 with BD has not been previously tested, meta-analyses of our associations in *NRG1* with those from previous reports in Turkish and Japanese [2, 3] indicate that the susceptibility conferred by these variants is not restricted to the Iranian population and strengthen the role of *NRG1* in BD susceptibility. The effect size of these *NRG1* variants in BD susceptibility (ORs for the risk alleles of approximately 1.20) is much smaller than the effect of the well-established *HLA-B51* allele (OR of approximately 3.50) [2, 3] but is of the same magnitude as reported non-HLA allelic associations (ORs typically between 1.20 and 1.60) in genome-wide associations studies for BD [2, 3, 17, 18]. Moreover, these three genetic markers have minor allele frequencies ranging from 0.122 to 0.473 in HapMap Caucasian (CEU) and Asian (CHB and JPT) populations, suggesting that these common variants may influence risk for BD in a substantial proportion of the population.

Concerns with multiple testing arise when screening a large number of genes or genetic variants. The results of the gene expression profiling were assessed for false discovery rate using the *Q* value, but no correction for the number of SNPs tested was performed. However, the *NRG1* meta-analyses and the observed epistatic

Table 3 Principal demographic and clinical characteristics of the Iranian case–control sample used in the association study

Characteristic	Controls	BD cases
N	839	976
Gender (*n*/*N*, % males)	348/839 (41.5)	512/976 (52.5)
Mean age-at-examination (years)±SD	40.0±12.3	39.1±11.0
Mean age-at-diagnosis (years)±SD	–	32.2±9.2
Oral aphthosis (*n*/*N*, %)	0/839 (0)	963/976 (98.7)
Genital aphthosis (*n*/*N*, %)	0/839 (0)	608/976 (62.3)
Skin lesions (*n*/*N*, %)	–	539/976 (55.2)
Pseudofolliculitis (*n*/*N*, %)	–	409/539 (75.9)
Erythema nodosum (*n*/*N*, %)	–	211/539 (39.1)
Skin aphthosis (*n*/*N*, %)	–	30/539 (5.6)
Ophthalmologic manifestations (*n*/*N*, %)	–	583/976 (59.7)
Anterior uveitis (*n*/*N*, %)	–	420/583 (72.0)
Posterior uveitis (*n*/*N*, %)	–	495/583 (84.9)
Retinal vasculitis (*n*/*N*, %)	–	343/583 (58.8)
Joint manifestations (*n*/*N*, %)	–	301/976 (30.8)
Arthralgia (*n*/*N*, %)	–	139/301 (46.2)
Arthritis (*n*/*N*, %)	–	188/301 (62.5)
Ankylosing spondylitis (*n*/*N*, %)	–	18/301 (6.0)
Neurological manifestations (*n*/*N*, %)	–	62/976 (6.4)
Vascular involvement (*n*/*N*, %)	–	51/976 (5.2)
Gastro-intestinal manifestations (*n*/*N*, %)	–	39/976 (4.0)
Epididymitis (*n*/*N*, %)	–	22/512 (4.3)
Cardiac involvement (*n*/*N*, %)	–	6/976 (0.6)
Pleuro-pulmonary involvement (*n*/*N*, %)	–	6/976 (0.6)
Pathergy phenomenon (*n*/*N*, %)	–	441/960 (45.9)
Family history of BD (*n*/*N*, %)	–	86/936 (9.2)

interactions between genetic markers in *NRG1* and *EREG-AREG* reinforce the role of this pathway in BD pathogenesis and warrant further validation in independent population samples.

To date, these four genetic markers have not been associated with other diseases (NHGRI catalogue of GWAS), and no functions have been linked to these variants. However, *NRG1* is encoded by 21 alternatively spliced exons spanning more than 1.1 Mb which, through alternative promoter usage and splicing, produce a variety of isoforms with different affinities to their receptors [19]. Interestingly, rs4489285 is in complete LD with rs1476540 (r^2=1, CEU population) that is located in a conserved transcription factor binding site for GATA-1, GATA-2, and GATA-3 in humans, and rs383632 is in complete LD with rs967205 (r^2=1, CEU population) which localizes to a conserved transcription factor binding site for FOXO3a, FOXO3b, and FOXD1 in humans (http://snp-nexus.org). The associated SNPs may therefore work as proxies for variants in intronic enhancers.

Epiregulin, the most under-expressed gene in BD patients, plays an essential role in immune/inflammatory-related responses in keratinocytes and macrophages in the epidermal layer [20] and has a pivotal role in peptidoglycan-induced proinflammatory cytokine production by antigen presenting cells [21]. Indeed, deficiency of epiregulin in mice results in chronic dermatitis and is correlated with an enhanced expression of the pro-inflammatory cytokine IL-18 by keratinocytes, supporting the role of this molecule in inflammatory diseases. An in silico analysis of the functional relevance of rs6845297 (located downstream of *EREG*) using SNPnexus (http://snp-nexus.org/) did not reveal previous associations of this polymorphism with other phenotypes or potential effects on regulatory elements or conserved sequences.

Amphiregulin binds exclusively to ErbB1 [22], and its overexpression has been associated with several autoimmune disorders such as systemic lupus erythematosus, psoriasis, and rheumatoid arthritis, and to synovial membrane inflammation [23–25]. Studies in transgenic mice demonstrate that amphiregulin overexpression in both the epidermis basal and suprabasal layer leads to a severe psoriasis-like phenotype and skin inflammation with a rich dermal and epidermal infiltration of neutrophils and lymphocytes [26]. Epidermal *AREG* expression is a possible mediator of innate cutaneous immunity and epidermal proliferation and a potential trigger of both cutaneous psoriasis and psoriatic arthritis [26].

Table 4 Association results of three SNPs in *NRG1* in the Iranian dataset, in the discovery GWAS samples, and in the overall meta-analyses

SNP	Gene	Allele	Dataset	*N*		Frequency		*P*	OR [95 % CI]
				Cases	Controls	Cases	Controls		
rs4489285	*NRG1*	G	Iranian	976	839	0.302	0.339	**2.01E-02**	0.85 [0.73–0.97]
			Japanese [Ref. 3]	611	737	0.123	0.151	**4.32E-02**	0.79 [0.63–0.99]
			Overall					**2.10E-03**	0.83 [0.74–0.93]
rs383632	*NRG1*	T	Iranian	976	839	0.180	0.153	**2.78E-02**	1.22 [1.02–1.46]
			Turkish [Ref. 2]	1211	1270	0.166	0.140	**9.16E-03**	1.23 [1.05–1.43]
			Overall					**6.61E-04**	1.23 [1.09–1.38]
rs1462891	*NRG1*	T	Iranian	976	839	0.261	0.231	**3.92E-02**	1.18 [1.01–1.37]
			Turkish [Ref. 2]	1215	1278	0.236	0.205	**7.86E-03**	1.20 [1.05–1.37]
			Overall					**7.88E-04**	1.19 [1.07–1.32]

At each marker, the OR and CI refer to the allele on the forward strand of the human genome reference sequence. Significant *P* values are bolded

Because of its multiple isoforms and various functional roles, *NRG1* has been implicated in the pathophysiology of numerous disorders, including breast cancer, schizophrenia, and bipolar disorder in human studies, atherosclerosis, myocardial dysfunction, and multiple sclerosis in animal models [27]. Interestingly, a schizophrenia-associated missense mutation in *NRG1* has been associated with immune system deregulation [27], establishing a link between this gene and immunological disorders. NRG1 can bind either to ErbB3 (that has no active kinase domain) and/or ErbB4. *ErbB4*, as well as two other genes of further downstream in the pathway (*DCN*, *SOS*) (Supplementary Table 1), were differentially expressed but not among the top genes, and thus were not selected for follow-up in the association study. On the other hand, *ErbB1* was not differentially expressed, suggesting that the intracellular signaling pathway affected in BD may be via ErbB4 and not ErbB1. To further establish the role of this pathway in BD pathogenesis, comprehensive immunogenetic/cellular studies may be conducted.

We cannot formally exclude the possibility that the relatively small number of samples and the different immunossupressants (IS) taken by some patients had an impact on our findings in the expression study, and may prevented the detection of some important BD genes. Nevertheless, to minimize the effect of this potential confounding factor, we included the immunosuppression status as a co-variate in the analysis. The rationale behind including immunosuppressed patients in the expression profiling was that they are expected to be in a physiological state that more closely resembles the inactive phase of BD patients not taking immunosuppressants drugs. Our group of cases is therefore thought to be more homogeneous than a group of patients with varying degrees of disease activity. Furthermore, current treatments for BD target mostly environmental triggers (e.g., antibiotics to treat bacterial infections) or self-amplifying disease mechanisms (e.g., immunosuppressants to block auto-immunity, NSAIDs and corticosteroids to reduce inflammation), but are not specific for this disorder and are unlikely to affect the unique initial etiopathogenic mechanism(s) that we are searching for. Also, unlike canonical pathways related to immunity and inflammation, the neuroregulin signaling pathway is not as likely to be influenced by immunomodulatory drugs.

In summary, we have shown the power of combining genome-wide expression studies with in silico pathways

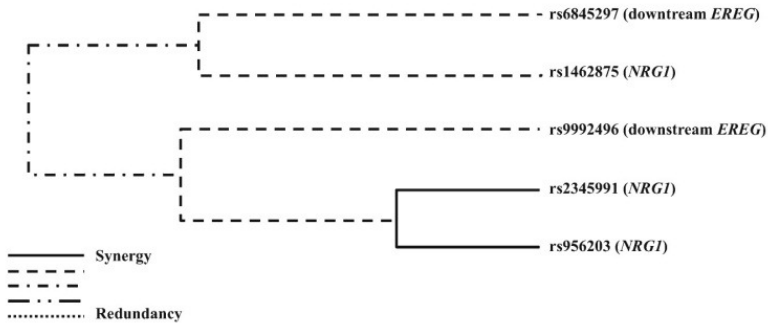


Fig. 2 Interaction dendrogram for the *EREG-AREG* and *NRG1* polymorphisms in BD susceptibility. The length of the dendrogram branch that connects two polymorphisms indicates the strength of interaction (the shorter the branch, the stronger is the interaction)

analysis and association studies in the identification of novel genetic risk factors for BD. We have also yielded a series of candidate genes, blood biomarkers, pathways, and mechanisms that are prime targets for follow-up in hypothesis driven studies. Future research towards understanding the role of these genes and pathways in BD pathogenesis include candidate gene association studies, deep re-sequencing of selected genes, biochemical, cellular, and animal work.

Acknowledgments We thank Drs. Ahmet Gul, Elaine F. Remmers, Shigeaki Ohno, Nobuhisa Mizuki and Akira Meguro for sharing their GWAS data. We thank to Dr. Majid Zeidi (Iranian Blood Transfusion Organization) for his excellent support. We are thankful to Doctor Sirous Zeinali (Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran) and Doctor Kayvan Saeedfar for their valuable help. We are also deeply grateful to all study participants and to the genotyping unit at the Instituto Gulbenkian de Ciência.

Funding This research was supported by the Research Committee of the Tehran University of Medical Sciences (grant 132/714), the Portuguese Fundação para a Ciência e a Tecnologia (grant PTDC/SAU-GMG/098937/2008, doctoral fellowship SFRH/BD/43895/2008 to JMX, and a Ciência contract to SAO), and the Portuguese Instituto do Emprego e Formação Profissional (fellowship to JMX, TK, BVF).

Disclosure of potential conflict of interest There is no conflict of interest to disclose.

References

1. Kapsimali VD, Kanakis MA, Vaiopoulos GA, Kaklamanis PG (2010) Etiopathogenesis of Behçet's disease with emphasis on the role of immunological aberrations. Clin Rheumatol 29:1211–1216
2. Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, Le JM, Yang B, Korman BD, Cakiris A et al (2010) Genome-wide association study identifies variants in the MHC class I, *IL10*, and *IL23R-IL12RB2* regions associated with Behçet's disease. Nat Genet 42:698–702
3. Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, Ito N, Kera J, Okada E, Yatsu K et al (2010) Genome-wide association studies identify *IL23R-IL12RB2* and *IL10* as Behçet's disease susceptibility loci. Nat Genet 42:703–706
4. Olsen N, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J, Aune TM (2004) A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. Ann Rheum Dis 63:1387–1392
5. Achiron A, Gurevich M, Friedman N, Kaminski N, Mandel M (2004) Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. Ann Neurol 55:410–417
6. Mandel M, Gurevich M, Pauzner R, Kaminski N, Achiron A (2004) Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. Clin Exp Immunol 138:164–170
7. International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD) (2006) Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 24(Suppl 42):S14–S15
8. Quian HR, Huang S (2005) Comparison of false discovery rate methods in identifying genes with differential expression. Genomics 86:495–503
9. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC et al (2001) Minimum information about a microarray experiment (MIAME): toward standards for microarray data. Nat Genet 29:365–371
10. Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, Abdollahi BS, Nadji A, Jesus G, Barcelos F, Vaz Patto J et al (2012) Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients. Arthritis Rheum 64:2761–2772
11. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
12. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644–645
13. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138–147
14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575
15. Moore JH (2004) Computational analysis of gene–gene interactions using multifactor dimensionality reduction. Expert Rev Mol Diagn 4:795–803
16. Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, Faezi T, Ghodsi Z, Faridar A, Ashofteh F et al (2010) Behçet's disease: from east to west. Clin Rheumatol 29:823–833
17. Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, Li F, Zhou Q, Ohno S, Chen R et al (2012) Identification of a susceptibility locus in *STAT4* for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum 64:4104–4013
18. Lee YJ, Horie Y, Wallace GR, Choi YS, Park JA, Song R, Kang YM, Kang SW, Baek HJ, Kitaichi N et al (2013) Genome-wide association study identifies *GIMAP* as a novel susceptibility locus for Behcet's disease. Ann Rheum Dis. doi:10.1136/annrheumdis-2011-200288
19. Falls DL (2003) Neuregulins: functions, forms, and signaling strategies. Exp Cell Res 284:14–30
20. Shirasawa S, Sugiyama S, Baba I, Inokuchi J, Sekine S, Ogino K, Kawamura Y, Dohi T, Fujimoto M, Sasazuki T (2004) Dermatitis due to epiregulin deficiency and a critical role of epiregulin in immune-related responses of keratinocyte and macrophage. Proc Natl Acad Sci U S A 101:13921–13926
21. Sugiyama S, Nakabayashi K, Baba I, Sasazuki T, Shirasawa S (2005) Role of epiregulin in peptidoglycan-induced proinflammatory cytokine production by antigen presenting cells. Biochem Biophys Res Commun 337:271–274
22. Shoyab M, Plowman GD, McDonald VL, Bradley JG, Todaro GJ (1989) Structure and function of human amphiregulin: a member of the epidermal growth factor family. Science 243:1074–1076
23. Ishii T, Onda H, Tanigawa A, Ohshima S, Fujiwara H, Mima T, Katada Y, Deguchi H, Suemura M, Miyake T et al (2005) Isolation and expression profiling of genes upregulated in the peripheral blood cells of systemic lupus erythematosus patients. DNA Res 12:429–439
24. Cook PW, Pittelkow MR, Keeble WW, Graves-Deal R, Coffey RJ Jr, Shipley GD (1992) Amphiregulin messenger RNA is elevated in psoriatic epidermis and gastrointestinal carcinomas. Cancer Res 52:3224–3227

25. Davies MR, Harding CJ, Raines S, Tolley K, Parker AE, Downey-Jones M, Needham MR (2005) Nurr1 dependent regulation of pro-inflammatory mediators in immortalised synovial fibroblasts. J Inflamm 2:15
26. Cook PW, Brown JR, Cornell KA, Pittelkow MR (2004) Suprabasal expression of human amphiregulin in the epidermis of transgenic mice induces a severe, earlyonset, psoriasis-like skin pathology: expression of amphiregulin in the basal epidermis is also associated with synovitis. Exp Dermatol 13:347–356
27. Marballi K, Quinones MP, Jimenez F, Escamilla MA, Raventós H, Soto-Bernardini MC, Ahuja SS, Walss-Bass C (2010) In vivo and in vitro genetic evidence of involvement of neuregulin 1 in immune system dysregulation. J Mol Med 88:1133–1141

5.2 SUPPLEMENTARY MATERIAL

Supplementary Table 1. Primer sequences used to genotype the 27 SNPs investigated in this study.

SNP ID	dbSNP ID	PCR primer 1	PCR primer 2	Extension primer
1	rs7687621	ACGTTGGATGGGGACAAAATAAACTCAAA	ACGTTGGATGAGGTCTCATAGCTAGGATTC	agaATAGCTAGGATTCATAGATTTATC
2	rs2061509	ACGTTGGATGGAAATTCCAAGAGGCGTGTG	ACGTTGGATGATTCCTTAGAAGGGAGGAAG	AGAGACAGAAAAAGAAAACCTAAAA
3	rs6816797	ACGTTGGATGGAGGTACCTGACAACACTTC	ACGTTGGATGCTATATCAAAGTGTGTGTCG	taatTGTGTGTCGAAACAGA
4	rs6845297	ACGTTGGATGGGGAAATTGAGACACAGGG	ACGTTGGATGAGACACCCTGAAATTGTAGC	cccTTTTATCACTAACCATGAGAC
5	rs12641287	ACGTTGGATGCAAGGGTAGAAACTTTAGGG	ACGTTGGATGCTCATTGTTCAATTCCCACC	TTCCCACCTATGTTCTTAT
6	rs9992496	ACGTTGGATGAGCACCTTGAGCATGTTACC	ACGTTGGATGAGTAGCCCTGTAAGGCAGAG	agggGCCCTGTAAGGCAGAGTTGCTA
7	rs1494882	ACGTTGGATGAGTTCTGAAAGAATGTTCC	ACGTTGGATGTGGTTTGAAGGATTAACTC	TGAAGGATTAACTCCTTTCA
8	rs7693713	ACGTTGGATGTAATAACAGCGTCCTCCCTC	ACGTTGGATGAATCAAGCCCCTTAGTGTGG	aTAGTGTGGCTACGGA
9	rs12503501	ACGTTGGATGCAAAGCTTCCAGCAAGTACC	ACGTTGGATGATGGGACTCTTTCCATGCTC	GGATAGTTTCTTGATGGC
10	rs1691273	ACGTTGGATGGGTTGATATATCATTGTCTC	ACGTTGGATGAGCCACAAAATTTCCTCATC	aTCCTAAAATACAACTGAGCTCTTA
11	rs1691274	ACGTTGGATGGTTAGTAAGTGTTATGAGAG	ACGTTGGATGTATCCCCTCGTCTTTTTGCC	CTTTTTGCCCCCAGG
12	rs4350980	ACGTTGGATGGGTCCAGTAATAGTAACCAAG	ACGTTGGATGACTGCCTTCTGGTAACTATG	gtgAGTAGCAATAGGGAAACAA
13	rs4489285	ACGTTGGATGCTCACCTGTAACAAAGTTGC	ACGTTGGATGGGTTTGGAGAGATGAGAATG	cccccGAGAGATGAGAATGAGAAGATGA
14	rs1557800	ACGTTGGATGAGAGATACATTTTCCCACAG	ACGTTGGATGCTATAAAATCATTCTTTCCCC	cATCATTCTTTCCCCTTTATTTAC
15	rs10954811	ACGTTGGATGCTAACTCAGAGTCAACATGG	ACGTTGGATGAGGTCTCAGAAAAGGTATTG	cttaAAAGAGAACAGGAAAGATGAT
16	rs2345991	ACGTTGGATGATCACTTTTGGAAAAGCTG	ACGTTGGATGTGAAACCTTCCTTCACTCCC	caCCTACCCTCTGACCTA
17	rs1503486	ACGTTGGATGTTGGATCTGGACTCTAGAAG	ACGTTGGATGCTGTTGTAGTTCTCCACCTC	gTCAATGCTCATTCCCGA
18	rs4733272	ACGTTGGATGCTTAAAGGGCTATGTGAGGG	ACGTTGGATGCATGTCTATACCTGAGACCC	gtaacGTCTTGATACCCCAACAT
19	rs16878317	ACGTTGGATGAATGACTGCTTGTTTTCACC	ACGTTGGATGGCGAGGTAACATCTGATGTG	AGCAAACAGTTGTTATTTTTACT
20	rs776385	ACGTTGGATGTTGCAGGTGAGTGGTCATCC	ACGTTGGATGTCCATTGGTGACAAGTAAGG	TACTCAAGGGGCACG
21	rs383632	ACGTTGGATGTGGCTTAGATCACTCTGAAG	ACGTTGGATGCTCCTGCCAAATGCTGTTGA	aatgGCTGTTGAAATTGCTGGGCTGGT
22	rs1462891	ACGTTGGATGGAATAAATCCCAGGCTTCAG	ACGTTGGATGGTTCCTTTGTATAAACTCG	cTATTTATTGAGCACCTACCA
23	rs17683983	ACGTTGGATGTTGGAATGTGATGTTCCTGC	ACGTTGGATGGCCTAAATTATACAACAGGA	gggaAGGATCCATGGGAGAA

24	rs1462875	ACGTTGGATGCAACCAGTTCTTCAAGGAGC	ACGTTGGATGAAGGGCACCCCAATAGTAAC	CCAATAGTAACAATCACACT
25	rs956203	ACGTTGGATGCCGCTAAGTCTCCATATTTG	ACGTTGGATGGTATTCTGACTAGCCCTGAG	ATGGACACAGCCCTC
26	rs16878764	ACGTTGGATGTCTTTCCATATGCCTGTCCC	ACGTTGGATGGACTAACATTTTTCCTGTTGC	TTTTCCTGTTGCTTAGAAC
27	rs10503899	ACGTTGGATGAAGTAATAGGCTCCTCCCAC	ACGTTGGATGCAGCAAAGTGACAGAAACGC	aaacGTGACAGAAACGCTCATAC

Supplementary Table 2. 621 probe sets representing 508 genes found differentially expressed among BD cases and controls with a threshold of 1.2 fold-change and $P \leq 0.05$.

Probe set ID	Gene symbol	Gene name	Cytoband	*P*-value	Fold change
204419_x_at	HBG1/HBG2	Hemoglobin, gamma A/ hemoglobin, gamma G	11p15.5	1.35E-02	2.53
206108_s_at	SRSF6	Serine/arginine-rich splicing factor 6	20q12-q13.1	3.96E-02	2.45
204848_x_at	HBG1/HBG2	Hemoglobin, gamma A/ hemoglobin, gamma G	11p15.5	1.29E-02	2.29
208450_at	LGALS2	Lectin, galactoside-binding, soluble, 2	22q12-q13|22q13.1	1.11E-02	2.20
203961_at	NEBL	Nebulette	10p12	2.59E-02	2.10
213515_x_at	HBG1/HBG2	Hemoglobin, gamma A/ hemoglobin, gamma G	11p15.5	2.03E-02	2.05
239287_at	---	---	---	4.54E-03	1.97
203962_s_at	NEBL	Nebulette	10p12	3.38E-02	1.87
234964_at	TRD@	T cell receptor delta locus	14q11.2	1.91E-02	1.81
222027_at	NUCKS1	Nuclear casein kinase and cyclin-dependent kinase substrate 1	1q32.1	3.06E-02	1.68
242577_at	LOC389834	Ankyrin repeat domain 57 pseudogene	---	2.18E-04	1.63
224009_x_at	DHRS9	Dehydrogenase/reductase member 9	2q31.1	4.92E-03	1.60
209642_at	BUB1	Budding uninhibited by benzimidazoles 1 homolog	2q14	4.54E-02	1.59
212671_s_at	HLA-DQA1/HLA-DQA2	Major histocompatibility complex (MHC), class II, DQ alpha 1/MCH, class II, DQ alpha 2	6p21.3	3.19E-02	1.57
216733_s_at	GATM	Glycine amidinotransferase	15q21.1	8.85E-03	1.55
221530_s_at	BHLHE41	Basic helix-loop-helix family, member e41	12p12.1	4.08E-03	1.53
223952_x_at	DHRS9	Dehydrogenase/reductase member 9	2q31.1	6.48E-03	1.52
223751_x_at	TLR10	Toll-like receptor 10	4p14	2.58E-03	1.51
206206_at	CD180	CD180 molecule	5q12	3.86E-03	1.51
204836_at	GLDC	Glycine dehydrogenase	9p22	3.30E-02	1.51
211634_x_at	IGHM/LOC100133862	Immunoglobulin (Ig) heavy constant mu/ Ig heavy chain V-I region V35-like	14q32.33	1.19E-02	1.50
216984_x_at	IGLV2-23	Immunoglobulin lambda variable 2-23	22q11.2	1.97E-02	1.49
242517_at	KISS1R	KISS1 receptor	19p13.3	2.90E-02	1.48

211635_x_at	IGHA1/ IGHA2/ IGHD /IGHG1/ IGHG3/ IGHG4/ IGHM/ IGHV4-31/ LOC100133862	immunoglobulin (Ig) heavy constant alpha 1/ Ig heavy constant alpha 2 / Ig heavy constant delta/ Ig heavy constant gamma 1 / Ig heavy constant gamma 3 /Ig heavy constant gamma 4 /Ig heavy constant mu/ Ig heavy variable 4-31/ Ig heavy chain V-I region V35-like	14q32.33	1.16E-02	1.47
219799_s_at	DHRS9	Dehydrogenase/reductase member 9	2q31.1	9.98E-03	1.46
225188_at	RAPH1	Ras association and pleckstrin homology domains 1	2q33	1.16E-02	1.46
212097_at	CAV1	Caveolin 1, caveolae protein, 22kDa	7q31.1	3.69E-02	1.44
225809_at	PARM1	Prostate androgen-regulated mucin-like protein 1	4q13.3-q21.3	3.26E-02	1.41
211655_at	LOC100287927	Hypothetical protein LOC100287927	22q11.22	1.59E-02	1.40
212390_at	PDE4DIP	Phosphodiesterase 4D interacting protein	1q12	1.85E-02	1.39
223750_s_at	TLR10	Toll-like receptor 10	4p14	4.81E-02	1.38
219146_at	C17orf42	Chromosome 17 open reading frame 42	---	1.14E-03	1.38
216133_at	TRD@	T cell receptor delta locus	14q11.2	4.78E-02	1.38
203178_at	GATM	Glycine amidinotransferase	15q21.1	4.65E-02	1.37
1558438_a_at	IGHE	Immunoglobulin heavy constant epsilon	14q32.33	6.56E-03	1.37
201841_s_at	HSPB1	Heat shock 27kDa protein 1	7q11.23	2.34E-02	1.34
224404_s_at	FCRL5	Fc receptor-like 5	1q21	1.99E-02	1.34
230879_at	BAG2	BCL2-associated athanogene 2	6p12.1-p11.2	6.87E-03	1.34
211868_x_at	IGHA1/ IGHA2 /IGHD /IGHG1 /IGHG2/ IGHG3 /IGHM /IGHV3-48 /IGHV4-31 /LOC100126583 /LOC100291917	Immunoglobulin (Ig) heavy constant alpha 1/ Ig heavy constant alpha 2 / Ig heavy constant delta/ Ig heavy constant gamma 1/ Ig heavy constant gamma 2 / Ig heavy constant gamma 3/Ig heavy constant mu/ Ig heavy variable 3-48/ Ig heavy variable 4-31/ hypothetical LOC100126583/ hypothetical protein LOC100291917	14q32.33	2.07E-02	1.33
216541_x_at	IGHG1/ IGHM/ LOC100133862	Immunoglobulin (Ig) heavy constant gamma 1/ Ig heavy constant mu/ Ig heavy chain V-I region V35-like	14q32.33	2.43E-02	1.33
1554345_a_at	GIN1	Gypsy retrotransposon integrase 1	5q21.1	3.16E-02	1.33
1558176_at	IGLC7/ IGLV1-44/ IGLV7-43/ LOC100290481	immunoglobulin (Ig) lambda constant 7/ Ig lambda variable 1-44/ Ig lambda variable 7-43/ Ig lambda light chain-like	22q11.2	2.54E-02	1.33
229779_at	COL4A4	Collagen, type IV, alpha 4	2q35-q37	2.35E-02	1.33
222284_at	---	---	---	7.27E-03	1.33

205987_at	CD1C	CD1c molecule	1q22-q23	1.85E-02	1.33
232975_at	HCG18	HLA complex group 18	6p21.3	1.35E-02	1.32
200983_x_at	CD59	CD59 molecule, complement regulatory protein	11p13	2.79E-02	1.32
237953_at	---	---	---	3.02E-02	1.32
1569318_at	LOC284440	Hypothetical LOC284440	19p13.11	5.27E-04	1.32
209660_at	TTR	Transthyretin	18q12.1	1.85E-02	1.32
212023_s_at	MKI67	Antigen identified by monoclonal antibody Ki-67	10q26.2	3.77E-04	1.31
204401_at	KCNN4	Potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4	19q13.2	3.08E-02	1.31
1556420_s_at	YPEL2	Yippee-like 2	17q22	1.86E-02	1.31
224406_s_at	FCRL5	Fc receptor-like 5	1q21	4.89E-02	1.31
1560109_s_at	---	---	---	8.53E-04	1.31
212952_at	LOC100507328	Hypothetical LOC100507328	---	3.74E-02	1.30
227458_at	CD274	CD274 molecule	9p24	3.41E-02	1.30
238454_at	ZNF540	Zinc finger protein 540	19q13.12	2.82E-02	1.30
231118_at	ANKRD35	Ankyrin repeat domain 35	1q21.1	2.14E-03	1.30
210395_x_at	MYL4	Myosin, light chain 4, alkali; atrial, embryonic	17q21-qter	3.46E-02	1.30
232239_at	LOC643529	hCG2024094	10q23.31	4.80E-02	1.30
237199_at	---	---	---	1.96E-02	1.30
200770_s_at	LAMC1	Laminin, gamma 1	1q31	1.84E-02	1.30
212021_s_at	MKI67	Antigen identified by monoclonal antibody Ki-67	10q26.2	4.87E-02	1.29
202490_at	IKBKAP	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	9q31	2.51E-02	1.29
220328_at	PHC3	Polyhomeotic homolog 3	3q26.2	5.83E-04	1.29
231176_at	PRR19	Proline rich 19	19q13.2	4.97E-03	1.29
228215_at	---	---	---	1.37E-03	1.29
213069_at	HEG1	HEG homolog 1	3q21.2	1.76E-03	1.29
214193_s_at	C1orf107	Chromosome 1 open reading frame 107	1q32.2	1.10E-02	1.28
205393_s_at	CHEK1	CHK1 checkpoint homolog	11q24.2	3.75E-02	1.28

203968_s_at	CDC6	Cell division cycle 6 homolog	17q21.3	2.98E-02	1.28
238598_s_at	---	---	---	1.39E-02	1.28
214933_at	CACNA1A	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	19p13	1.08E-02	1.28
214460_at	LSAMP	Limbic system-associated membrane protein	3q13.2-q21	1.38E-02	1.28
203711_s_at	HIBCH	3-hydroxyisobutyryl-CoA hydrolase	2q32.2	3.64E-02	1.28
237709_at	LOC100507513	Hypothetical LOC100507513	---	1.28E-03	1.28
237524_at	---	---	---	1.35E-04	1.28
230729_at	---	---	---	4.14E-03	1.28
243352_at	ALPK1	Alpha-kinase 1	4q25	2.81E-02	1.27
224797_at	ARRDC3	Arrestin domain containing 3	5q14.3	4.01E-03	1.27
235867_at	GSTM3	Glutathione S-transferase mu 3	1p13.3	3.17E-02	1.27
230359_at	KNDC1	Kinase non-catalytic C-lobe domain containing 1	10q26.3	1.77E-02	1.27
209986_at	ASCL1	Achaete-scute complex homolog 1	12q23.2	6.56E-03	1.27
233253_at	---	---	---	5.60E-03	1.27
204033_at	TRIP13	Thyroid hormone receptor interactor 13	5p15.33	4.52E-02	1.27
230795_at	---	---	---	3.35E-02	1.27
238071_at	LCN10/LCN6	Lipocalin 10/lipocalin 6	9q34.3	6.63E-03	1.27
211640_x_at	IGHG1/ IGHM/ LOC100133862	Immunoglobulin (Ig) heavy constant gamma 1/ Ig heavy constant mu/Ig heavy chain V-I region V35-like	14q32.33	2.41E-02	1.27
226250_at	---	---	---	2.27E-02	1.27
243010_at	MSI2	Musashi homolog 2	17q22	3.36E-03	1.27
231428_at	---	---	---	3.48E-02	1.27
203144_s_at	KIAA0040	KIAA0040	1q24-q25	3.87E-02	1.27
214342_at	ATXN7L1	Ataxin 7-like 1	7q22.3	6.52E-04	1.26
201129_at	SRSF7	Serine/arginine-rich splicing factor 7	2p22.1	1.20E-02	1.26
241888_at	---	---	---	4.64E-02	1.26
231771_at	GJB6	Gap junction protein, beta 6, 30kDa	13q11-q12.1|13q12	2.46E-02	1.26
239425_at	DCUN1D5	DCN1, defective in cullin neddylation 1, domain containing 5	11q22.3	4.59E-02	1.26

229559_at	PPM1N	Protein phosphatase, Mg2+/Mn2+ dependent, 1N	19q13.32	2.31E-02	1.26
1555042_at	TRPV5	Transient receptor potential cation channel, subfamily V, member 5	7q35	3.80E-03	1.26
241141_at	BMP6	Bone morphogenetic protein 6	6p24-p23	8.94E-03	1.25
232923_at	---	---	---	5.84E-04	1.25
228773_at	LOC100506100	Hypothetical LOC100506100	---	2.40E-02	1.25
234707_x_at	---	---	---	4.48E-02	1.25
225706_at	GLCCI1	Glucocorticoid induced transcript 1	7p21.3	2.04E-02	1.25
224364_s_at	PPIL3	Peptidylprolyl isomerase -like 3	2q33.1	2.73E-02	1.25
241732_at	---	---	---	4.13E-02	1.25
202759_s_at	AKAP2/ PALM2-AKAP2	A kinase anchor protein 2/PALM2-AKAP2 readthrough	9q31.3	3.62E-02	1.25
238639_x_at	LOC100288271	Hypothetical LOC100288271	1p36.33	1.14E-04	1.25
204201_s_at	PTPN13	Protein tyrosine phosphatase, non-receptor type 13	4q21.3	4.66E-02	1.25
239832_at	---	---	---	2.33E-02	1.25
230370_x_at	STYXL1	Serine/threonine/tyrosine interacting-like 1	7q11.23	8.00E-03	1.25
231374_at	---	---	---	1.15E-02	1.25
211718_at	MGC2889	Hypothetical protein MGC2889	3q29	4.75E-02	1.25
238032_at	---	---	---	4.18E-02	1.25
224324_at	MRO	Maestro	18q21	8.86E-03	1.25
236712_at	---	---	---	1.23E-02	1.25
243911_at	---	---	---	1.13E-02	1.25
206266_at	GPLD1	Glycosylphosphatidylinositol specific phospholipase D1	6p22.1	1.34E-02	1.25
204783_at	MLF1	Myeloid leukemia factor 1	3q25.1	4.40E-04	1.24
244327_at	---	---	---	5.73E-03	1.24
1553923_at	SLC22A24	Solute carrier family 22, member 24	11q12.3	4.52E-03	1.24
204720_s_at	DNAJC6	DnaJ homolog, subfamily C, member 6	1p31.3	3.21E-02	1.24
205783_at	KLK13	Kallikrein-related peptidase 13	19q13.33	2.68E-02	1.24
241841_at	---	---	---	9.51E-03	1.24
225123_at	---	---	---	4.18E-02	1.24

231114_at	SPATA22	Spermatogenesis associated 22	17p13.3	3.62E-03	1.24
217230_at	EZR	Ezrin	6q25.3	4.84E-02	1.24
243522_at	---	---	---	4.78E-02	1.24
231285_at	STT3B	STT3, subunit of the oligosaccharyltransferase complex, homolog B	3p23	5.06E-03	1.24
234677_at	---	---	---	3.53E-04	1.24
201893_x_at	DCN	Decorin	12q21.33	5.83E-03	1.24
214908_s_at	TRRAP	Transformation/transcription domain-associated protein	7q21.2-q22.1	2.46E-02	1.24
233982_x_at	STYXL1	Serine/threonine/tyrosine interacting-like 1	7q11.23	1.55E-02	1.24
207068_at	ZFP37	Zinc finger protein 37 homolog	9q32	1.10E-02	1.24
221238_at	HMGN5	High-mobility group nucleosome binding domain 5	Xq13.3	4.45E-02	1.24
1554094_at	ENTPD5	Ectonucleoside triphosphate diphosphohydrolase 5	14q24	2.10E-02	1.24
218321_x_at	STYXL1	Serine/threonine/tyrosine interacting-like 1	7q11.23	1.40E-02	1.24
223666_at	SNX5	Sorting nexin 5	20p11	4.21E-02	1.24
205553_s_at	CSRP3	Cysteine and glycine-rich protein 3	11p15.1	3.06E-03	1.23
237795_s_at	SP2	Sp2 transcription factor	17q21.32	8.12E-04	1.23
232851_at	FBXO3	F-box protein 3	11p13	1.58E-03	1.23
244851_at	---	---	---	1.39E-02	1.23
225931_s_at	RNF213	Ring finger protein 213	17q25.3	3.77E-02	1.23
232986_at	ZNF233	Zinc finger protein 233	19q13.31	3.58E-03	1.23
225700_at	GLCCI1	Glucocorticoid induced transcript 1	7p21.3	3.22E-02	1.23
228227_at	ITGB1BP1	Integrin beta 1 binding protein 1	2p25.2	2.30E-02	1.23
226383_at	C11orf46	Chromosome 11 open reading frame 46	11p14.1	3.87E-02	1.23
214641_at	COL4A3	Collagen, type IV, alpha 3	2q36-q37	8.20E-03	1.23
230194_at	LRPPRC	Leucine-rich PPR-motif containing	2p21	1.10E-02	1.23
205864_at	SLC7A4	Solute carrier family 7, member 4	22q11.21	1.74E-02	1.23
1569114_at	LOC100506847	Hypothetical LOC100506847	---	8.58E-03	1.23
237325_at	C22orf43	Chromosome 22 open reading frame 43	22q11.2	2.30E-02	1.23
228655_at	---	---	---	1.53E-02	1.23

216092_s_at	SLC7A8	Solute carrier family 7, member 8	14q11.2	4.53E-02	1.23
211438_at	TRHR	Thyrotropin-releasing hormone receptor	8q23	2.14E-02	1.23
240998_at	---	---	---	1.05E-02	1.23
229526_at	AQP11	Aquaporin 11	11q14.1	4.02E-03	1.23
224801_at	NDFIP2	Nedd4 family interacting protein 2	13q31.1	2.14E-02	1.23
207202_s_at	NR1I2	Nuclear receptor subfamily 1, group I, member 2	3q12-q13.3	5.69E-03	1.23
230730_at	SGCD	Sarcoglycan, delta	5q33-q34	2.71E-02	1.22
205248_at	DOPEY2	Dopey family member 2	21q22.2	2.48E-02	1.22
241515_at	---	---	---	3.02E-03	1.22
242689_at	RALGPS1	Ral GEF with PH domain and SH3 binding motif 1	9q33.3	1.25E-02	1.22
209353_s_at	---	---	---	4.90E-03	1.22
1555073_at	MGC40069	Hypothetical protein MGC40069	14q11.2	1.15E-02	1.22
221594_at	C7orf64	Chromosome 7 open reading frame 64	7q21.2	3.26E-02	1.22
203158_s_at	GLS	Glutaminase	2q32-q34	2.54E-02	1.22
229137_at	FUCA1	Fucosidase, alpha-L- 1, tissue	1p34	2.44E-02	1.22
209032_s_at	CADM1	Cell adhesion molecule 1	11q23.2	2.15E-03	1.22
222936_s_at	PPPDE1	PPPDE peptidase domain containing 1	1q44	1.89E-02	1.22
205991_s_at	PRRX1	Paired related homeobox 1	1q24	2.37E-03	1.22
241857_at	---	---	---	7.87E-03	1.22
1562712_at	---	---	---	4.52E-02	1.22
236776_at	---	---	---	4.34E-02	1.22
1557641_at	---	---	---	5.82E-03	1.22
233494_at	ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4	2q33.3-q34	5.56E-03	1.22
231945_at	FILIP1	Filamin A interacting protein 1	6q14.1	1.61E-02	1.22
205167_s_at	CDC25C	Cell division cycle 25 homolog C	5q31	1.02E-03	1.22
224530_s_at	KCNIP4	Kv channel interacting protein 4	4p15.32	8.74E-03	1.22
229928_at	LOC100507265	Hypothetical LOC100507265	---	2.59E-02	1.22
224360_s_at	PACSIN1	Protein kinase C and casein kinase substrate in neurons 1	6p21.3	4.24E-02	1.22

1568787_at	---	---	---	2.53E-02	1.22
213558_at	PCLO	Piccolo	7q11.23-q21.3	8.10E-05	1.22
220208_at	ADAMTS13	ADAM metallopeptidase with thrombospondin type 1 motif, 13	9q34	1.57E-02	1.22
1567251_at	OR10A3	Olfactory receptor, family 10, subfamily A, member 3	11p15.4	3.90E-02	1.22
229347_at	---	---	---	1.04E-02	1.22
216079_at	EPM2A	Epilepsy, progressive myoclonus type 2A	6q24	2.49E-02	1.22
211814_s_at	CCNE2	Cyclin E2	8q22.1	2.24E-02	1.22
225342_at	AK4	Adenylate kinase 4	1p31.3	4.27E-02	1.22
212822_at	HEG1	HEG homolog 1	3q21.2	4.07E-02	1.22
210673_x_at	NKX2-1	NK2 homeobox 1	14q13	3.85E-02	1.22
230093_at	RSPH1	Radial spoke head 1 homolog	21q22.3	1.22E-03	1.22
1568592_at	TRIM69	Tripartite motif-containing 69	15q21.1	1.23E-02	1.22
243408_at	---	---	---	2.60E-02	1.21
1560284_at	---	---	---	2.94E-02	1.21
223778_at	KIF9	Kinesin family member 9	3p21.31	6.03E-03	1.21
201408_at	PPP1CB	Protein phosphatase 1, catalytic subunit, beta isozyme	2p23	1.91E-02	1.21
208596_s_at	UGT1A1/ UGT1A10/ UGT1A3/ UGT1A4/ UGT1A5/ UGT1A6/ UGT1A7/ UGT1A8/ UGT1A9	UDP glucuronosyltransferase 1 family (UGT1), polypeptide A1/ UGT1, polypeptide A10/ UGT1, polypeptide A3/UGT1, polypeptide A4/ UGT1, polypeptide A5/ UGT1, polypeptide A6/ UGT1, polypeptide A7/ UGT1, polypeptide A8/ UGT1, polypeptide A9	2q37	4.24E-02	1.21
1561625_at	---	---	---	6.53E-03	1.21
1554681_a_at	MGC50722	Hypothetical MGC50722	---	3.26E-02	1.21
1564151_at	---	---	---	1.13E-02	1.21
223074_s_at	AIF1L	Allograft inflammatory factor 1-like	9q34.13-q34.3	3.19E-02	1.21
231656_x_at	OSBPL10	Oxysterol binding protein-like 10	3p22.3	4.73E-02	1.21
1563271_at	---	---	---	2.24E-02	1.21
224770_s_at	NAV1	Neuron navigator 1	1q32.3	3.07E-02	1.21
236480_at	---	---	---	1.37E-02	1.21

1569171_a_at	FXR1	Fragile X mental retardation, autosomal homolog 1	3q28	1.26E-03	1.21
209198_s_at	SYT11	Synaptotagmin XI	1q21.2	3.44E-02	1.21
204240_s_at	SMC2	Structural maintenance of chromosomes 2	9q31.1	3.24E-02	1.21
243515_at	---	---	---	1.36E-03	1.21
236378_at	CIB4	Calcium and integrin binding family member 4	2p23.3	4.78E-02	1.21
224449_at	DDI2	DNA-damage inducible 1 homolog 2	1p36.21	9.04E-03	1.21
242270_at	---	---	---	1.18E-02	1.21
231217_at	---	---	---	5.00E-02	1.21
201923_at	PRDX4	Peroxiredoxin 4	Xp22.11	3.75E-02	1.21
221296_at	TECTA	Tectorin alpha	11q22-q24	3.93E-02	1.21
1564548_at	LOC642620	hCG1999814	7q36.3	2.61E-02	1.21
1555377_at	OR4D2	Olfactory receptor, family 4, subfamily D, member 2	17q22	7.65E-03	1.21
235842_at	---	---	---	2.36E-02	1.21
201387_s_at	UCHL1	Ubiquitin carboxyl-terminal esterase L1	4p14	2.76E-03	1.21
228255_at	ALS2CR4	Amyotrophic lateral sclerosis 2 chromosome region, candidate 4	---	7.09E-03	1.21
213261_at	TRANK1	Tetratricopeptide repeat and ankyrin repeat containing 1	3p22.2	2.51E-02	1.21
210871_x_at	SSX2IP	Synovial sarcoma, X breakpoint 2 interacting protein	1p22.3	1.66E-02	1.21
244496_at	---	---	---	4.12E-03	1.21
208511_at	PTTG3P	Pituitary tumor-transforming 3	8q13.1	1.95E-03	1.21
209521_s_at	AMOT	Angiomotin	Xq23	4.03E-03	1.21
1569909_at	KRT79	Keratin 79	12q13.13	4.23E-02	1.21
233900_at	---	---	---	1.66E-02	1.21
1556195_a_at	---	---	---	2.39E-02	1.21
229314_at	---	---	---	3.57E-02	1.20
1552895_a_at	ANKRD30BP2	Ankyrin repeat domain 30B pseudogene 2	21q11.2	1.67E-03	1.20
233388_at	---	---	---	3.19E-03	1.20
223998_at	TTLL2	Tubulin tyrosine ligase-like family, member 2	6q27	1.60E-02	1.20
210051_at	RAPGEF3	Rap guanine nucleotide exchange factor 3	12q13.1	1.77E-02	1.20

217172_at	---	---	---	9.44E-03	1.20
232054_at	PCDH20	Protocadherin 20	13q21	1.66E-02	1.20
202190_at	CSTF1	Cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kDa	20q13.2	4.30E-02	1.20
209343_at	EFHD1	EF-hand domain family, member D1	2q37.1	3.84E-02	1.20
227311_at	SNX25	Sorting nexin 25	4q35.1	1.58E-02	1.20
228752_at	EFCAB4B	EF-hand calcium binding domain 4B	12p13.32	4.87E-02	1.20
207323_s_at	MBP	Myelin basic protein	18q23	3.50E-02	1.20
235153_at	RNF183	Ring finger protein 183	9q32	1.48E-02	1.20
233592_at	ANKRD18A/ ANKRD18B	Ankyrin repeat domain 18A/ ankyrin repeat domain 18B	9p13.1 / 9p13.3	1.20E-02	1.20
207142_at	KCNJ3	Potassium inwardly-rectifying channel, subfamily J, member 3	2q24.1	3.27E-02	1.20
1560002_at	FAM27D1	Family with sequence similarity 27, member D1	9p11.2	5.95E-03	1.20
1561911_at	---	---	---	1.41E-02	1.20
1554836_at	USP36	Ubiquitin specific peptidase 36	17q25.3	3.23E-02	1.20
208622_s_at	EZR	Ezrin	6q25.3	3.46E-02	1.20
229675_at	MINA	MYC induced nuclear antigen	3q11.2	9.47E-03	1.20
1556872_s_at	IQSEC3	IQ motif and Sec7 domain 3	12p13.33	2.83E-02	1.20
220255_at	FANCE	Fanconi anemia, complementation group E	6p22-p21	4.19E-02	1.20
1555158_at	---	---	---	5.20E-05	1.20
1561165_a_at	DEFB108B/ DEFB108P1	Defensin, beta 108B/ defensin, beta 108, pseudogene 1	11q13.4 / 8p23.1	1.99E-02	1.20
224499_s_at	AICDA	Activation-induced cytidine deaminase	12p13	3.96E-02	1.20
232353_s_at	STYXL1	Serine/threonine/tyrosine interacting-like 1	7q11.23	1.63E-02	1.20
1562446_at	ZNF391	Zinc finger protein 391	6p22.1	2.39E-02	1.20
221085_at	TNFSF15	Tumor necrosis factor superfamily, member 15	9q32	2.47E-02	1.20
228698_at	SOX7	SRY -box 7	8p22	1.47E-02	1.20
204368_at	SLCO2A1	Solute carrier organic anion transporter family, member 2A1	3q21	5.68E-03	1.20
219330_at	VANGL1	Vang-like 1	1p13.1	4.70E-02	-1.20
202047_s_at	CBX6	Chromobox homolog 6	22q13.1	8.34E-03	-1.20
200919_at	PHC2	Polyhomeotic homolog 2	1p34.3	2.52E-02	-1.20

229323_at	LOC387723	Hypothetical LOC387723	10q26.3	1.24E-02	-1.20
216629_at	SRRM2	Serine/arginine repetitive matrix 2	16p13.3	3.51E-02	-1.20
225825_at	C20orf194	Chromosome 20 open reading frame 194	20p13	2.45E-02	-1.20
203853_s_at	GAB2	GRB2-associated binding protein 2	11q14.1	7.70E-03	-1.20
219047_s_at	ZNF668	Zinc finger protein 668	16p11.2	1.47E-02	-1.20
214814_at	YTHDC1	YTH domain containing 1	4q13.2	2.62E-02	-1.20
226334_s_at	AHSA2	AHA1, activator of heat shock 90kDa protein ATPase homolog 2	2p15	3.64E-02	-1.20
228628_at	SRGAP2P1	SLIT-ROBO Rho GTPase activating protein 2 pseudogene 1	1p11.2	3.19E-02	-1.20
238038_at	---	---	---	3.73E-02	-1.20
220253_s_at	LRP12	Low density lipoprotein receptor-related protein 12	8q22.2	4.45E-02	-1.20
208458_at	SCNN1D	Sodium channel, nonvoltage-gated 1, delta	1p36.3-p36.2	2.91E-02	-1.20
240538_at	---	---	---	4.07E-02	-1.20
224563_at	WASF2	WAS protein family, member 2	1p36.11	4.86E-02	-1.20
244415_at	---	---	---	3.63E-02	-1.20
209889_at	SEC31B	SEC31 homolog B	10q24.31	3.18E-02	-1.20
1552272_a_at	PRR22	Proline rich 22	19p13.3	2.93E-02	-1.21
208015_at	SMAD1	SMAD family member 1	4q31	1.78E-02	-1.21
228662_at	LOC100505522	Hypothetical LOC100505522	---	1.02E-02	-1.21
237502_at	---	---	---	7.88E-03	-1.21
1556159_at	---	---	---	3.98E-02	-1.21
1568733_at	C10orf76	Chromosome 10 open reading frame 76	10q24.32	1.11E-02	-1.21
227106_at	LOC440104	1110012D08Rik pseudogene	12q13.2	5.46E-03	-1.21
218969_at	PAM16	Presequence translocase-associated motor 16 homolog	16p13.3	1.96E-02	-1.21
212262_at	QKI	Quaking homolog, KH domain RNA binding	6q26	4.54E-02	-1.21
210613_s_at	SYNGR1	Synaptogyrin 1	22q13.1	3.46E-02	-1.21
226691_at	TNRC18	Trinucleotide repeat containing 18	7p22.1	2.14E-02	-1.21
239795_at	---	---	---	1.92E-02	-1.21
214805_at	EIF4A1	Eukaryotic translation initiation factor 4A1	17p13	3.03E-02	-1.21

226644_at	MIB2	Mindbomb homolog 2	1p36.33	3.12E-02	-1.21
238153_at	PDE6B	Phosphodiesterase 6B, cGMP-specific, rod, beta	4p16.3	1.67E-02	-1.21
228957_at	ZNF362	Zinc finger protein 362	1p35.1	4.78E-02	-1.21
221590_s_at	ALDH6A1	Aldehyde dehydrogenase 6 family, member A1	14q24.3	3.77E-02	-1.21
214377_s_at	CTRL	Chymotrypsin-like	16q22.1	2.42E-02	-1.21
236023_at	CDK9	Cyclin-dependent kinase 9	9q34.1	2.85E-02	-1.21
241001_at	---	---	---	8.91E-03	-1.21
1567214_a_at	PNN	Pinin, desmosome associated protein	14q21.1	4.29E-02	-1.21
235554_x_at	PACRGL	PARK2 co-regulated-like	4p15.31	3.48E-02	-1.21
1570433_at	TMPRSS2	Transmembrane protease, serine 2	21q22.3	9.42E-03	-1.21
208032_s_at	GRIA3	Glutamate receptor, ionotrophic, AMPA 3	Xq25	1.32E-02	-1.21
243195_s_at	ZNF551	Zinc finger protein 551	19q13.43	3.15E-02	-1.21
211782_at	IDS	Iduronate 2-sulfatase	Xq28	8.25E-03	-1.21
229583_at	---	---	---	1.06E-02	-1.21
228196_s_at	LARP4B	La ribonucleoprotein domain family, member 4B	10p15.3	1.12E-02	-1.21
208317_at	XYLB	Xylulokinase homolog	3p22-p21.3	1.80E-02	-1.21
232306_at	CDH26	Cadherin 26	20q13.33	3.68E-02	-1.21
221566_s_at	NOL3	Nucleolar protein 3	16q22.1	2.96E-02	-1.22
229445_at	CYBA	Cytochrome b-245, alpha polypeptide	16q24	2.83E-02	-1.22
236381_s_at	WDR8	WD repeat domain 8	1p36.3	2.53E-02	-1.22
201926_s_at	CD55	CD55 molecule, decay accelerating factor for complement	1q32	1.82E-02	-1.22
230435_at	LOC375190	Hypothetical protein LOC375190	2p23.3	6.49E-03	-1.22
225249_at	SPPL2B	Signal peptide peptidase-like 2B	19p13.3	1.58E-02	-1.22
225818_s_at	TBRG1	Transforming growth factor beta regulator 1	11q24.2	2.71E-02	-1.22
235721_at	DTX3	Deltex homolog 3	12q13.3	1.77E-02	-1.22
230566_at	C22orf27	Chromosome 22 open reading frame 27	22q12.2	1.68E-02	-1.22
228983_at	---	---	---	3.68E-02	-1.22
218419_s_at	TMUB2	Transmembrane and ubiquitin-like domain containing 2	17q21.31	1.59E-02	-1.22

209528_s_at	TELO2	TEL2, telomere maintenance 2, homolog	16p13.3	7.91E-03	-1.22
221852_at	---	---	---	5.46E-03	-1.22
226597_at	REEP6	Receptor accessory protein 6	19p13.3	7.70E-03	-1.22
230995_at	CMBL	Carboxymethylenebutenolidase homolog	5p15.2	3.23E-02	-1.22
210230_at	---	---	---	4.47E-02	-1.22
228155_at	C10orf58	Chromosome 10 open reading frame 58	10q23.1	4.90E-02	-1.22
223276_at	C5orf62	Chromosome 5 open reading frame 62	5q33.1	1.39E-02	-1.22
231238_at	LOC100506075	Hypothetical LOC100506075	---	3.36E-02	-1.22
209696_at	FBP1	Fructose-1,6-bisphosphatase 1	9q22.3	4.15E-02	-1.22
210294_at	TAPBP	TAP binding protein	6p21.3	1.49E-02	-1.22
218821_at	NPEPL1	Aminopeptidase-like 1	20q13.32	2.21E-02	-1.22
1555972_s_at	FBXO28	F-box protein 28	1q42.11	3.66E-02	-1.22
202393_s_at	KLF10	Kruppel-like factor 10	8q22.2	4.49E-02	-1.22
212662_at	PVR	Poliovirus receptor	19q13.2	2.15E-02	-1.22
238980_x_at	C17orf56	Chromosome 17 open reading frame 56	17q25.3	2.24E-02	-1.22
202192_s_at	GAS7	Growth arrest-specific 7	17p13.1	4.54E-02	-1.22
232172_at	CD99P1	CD99 molecule pseudogene 1	Xp22.33;Yp11.31	1.41E-02	-1.22
241445_at	---	---	---	2.40E-02	-1.22
215114_at	SENP3	SUMO1/sentrin/SMT3 specific peptidase 3	17p13	4.67E-02	-1.22
230039_at	---	---	---	3.24E-02	-1.22
1555091_at	PPM1F	Protein phosphatase, Mg2+/Mn2+ dependent, 1F	22q11.22	4.66E-02	-1.22
226205_at	ANKRD13D	Ankyrin repeat domain 13 family, member D	11q13.2	2.74E-02	-1.23
219525_at	SLC47A1	Solute carrier family 47, member 1	17p11.2	1.92E-02	-1.23
226160_at	H6PD	Hexose-6-phosphate dehydrogenase	1p36	1.70E-02	-1.23
229431_at	RFXAP	Regulatory factor X-associated protein	13q14	3.34E-02	-1.23
219690_at	TMEM149	Transmembrane protein 149	19q13.12	3.31E-02	-1.23
202430_s_at	PLSCR1	Phospholipid scramblase 1	3q23	2.97E-02	-1.23
229890_at	PRRT1	Proline-rich transmembrane protein 1	6p21.32	3.01E-02	-1.23

221860_at	HNRNPL	Heterogeneous nuclear ribonucleoprotein L	19q13.2	1.21E-02	-1.23
238560_at	CALCOCO2	Calcium binding and coiled-coil domain 2	17q21.32	6.08E-03	-1.23
243680_at	LOC100506476	Hypothetical LOC100506476	---	2.07E-02	-1.23
238098_at	---	---	---	1.27E-02	-1.23
1560007_at	LOC645984	Hypothetical LOC645984	10p11.23	4.64E-02	-1.23
213935_at	ABHD5	Abhydrolase domain containing 5	3p21	4.72E-02	-1.23
55872_at	ZNF512B	Zinc finger protein 512B	20q13.33	3.50E-02	-1.23
212268_at	SERPINB1	Serpin peptidase inhibitor, clade B, member 1	6p25	1.35E-02	-1.23
243699_at	LOC100507006	Hypothetical LOC100507006	---	7.74E-03	-1.23
237276_at	NASP	Nuclear autoantigenic sperm protein	1p34.1	1.35E-02	-1.23
230221_at	BAT5	HLA-B associated transcript 5	6p21.3	1.39E-02	-1.23
218739_at	ABHD5	Abhydrolase domain containing 5	3p21	1.84E-02	-1.23
231200_at	LSM14B	LSM14B, SCD6 homolog B	20q13.33	3.27E-03	-1.23
202016_at	MEST	Mesoderm specific transcript homolog	7q32	4.26E-02	-1.23
240793_at	TTN	Titin	2q31	8.92E-03	-1.23
235140_at	SHROOM1	Shroom family member 1	5q31.1	3.54E-02	-1.23
227833_s_at	MBD6	Methyl-CpG binding domain protein 6	---	4.19E-03	-1.23
35776_at	ITSN1	Intersectin 1	21q22.1-q22.2	1.14E-02	-1.23
236969_at	---	---	---	1.52E-02	-1.24
206356_s_at	GNAL	Guanine nucleotide binding protein, alpha activating activity polypeptide, olfactory type	18p11.22-p11.21	4.02E-02	-1.24
209352_s_at	SIN3B	SIN3 homolog B, transcription regulator	19p13.11	4.48E-02	-1.24
208438_s_at	FGR	Gardner-Rasheed feline sarcoma viral oncogene homolog	1p36.2-p36.1	2.83E-02	-1.24
224558_s_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	11q13.1	2.71E-02	-1.24
219693_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	6q26	8.40E-03	-1.24
1555842_at	CYTH2	Cytohesin 2	19q13.33	3.20E-02	-1.24
236180_at	---	---	---	1.66E-02	-1.24
240260_at	---	---	---	4.94E-02	-1.24
227013_at	LATS2	LATS, large tumor suppressor, homolog 2	13q11-q12	1.27E-02	-1.24

1552323_s_at	FAM122C	Family with sequence similarity 122C	Xq26.3	4.10E-02	-1.24
228847_at	EXOC3	Exocyst complex component 3	5p15.33	1.54E-02	-1.24
228785_at	ZNF281	Zinc finger protein 281	1q32.1	4.20E-02	-1.24
209295_at	TNFRSF10B	Tumor necrosis factor receptor superfamily, member 10b	8p22-p21	1.01E-02	-1.24
240095_at	---	---	---	1.46E-02	-1.24
235889_at	---	---	---	4.61E-03	-1.24
232884_s_at	ZNF853	Zinc finger protein 853	7p22.1	2.06E-02	-1.24
218631_at	AVPI1	Arginine vasopressin-induced 1	10q24.2	4.94E-02	-1.24
218017_s_at	HGSNAT	Heparan-alpha-glucosaminide N-acetyltransferase	8p11.1	4.48E-02	-1.24
227542_at	SOCS6	Suppressor of cytokine signaling 6	18q22.2	4.07E-02	-1.24
237741_at	SLC25A36	Solute carrier family 25, member 36	3q23	1.36E-02	-1.24
242437_at	---	---	---	3.09E-02	-1.24
211840_s_at	PDE4D	Phosphodiesterase 4D, cAMP-specific	5q12	2.58E-02	-1.24
224829_at	CPEB4	Cytoplasmic polyadenylation element binding protein 4	5q21	3.97E-02	-1.24
241955_at	HECTD1	HECT domain containing 1	14q12	4.12E-02	-1.25
235888_at	GUSBP1	Glucuronidase, beta pseudogene 1	5p14.3	1.90E-02	-1.25
242442_x_at	RG9MTD2	RNA methyltransferase domain containing 2	4q23	1.02E-02	-1.25
220469_at	COPE	Coatomer protein complex, subunit epsilon	19p13.11	2.46E-02	-1.25
232915_at	DDX49	DEAD box polypeptide 49	19p12	5.09E-03	-1.25
223173_at	SPNS1	Spinster homolog 1	16p11.2	3.22E-02	-1.25
221698_s_at	CLEC7A	C-type lectin domain family 7, member A	12p13.2	9.69E-03	-1.25
1554182_at	TRIM73/ TRIM74	Tripartite motif-containing 73/ tripartite motif-containing 74	7q11.23	4.19E-02	-1.25
225640_at	LOC401504	Hypothetical LOC401504	9p13.2	3.99E-02	-1.25
237560_at	MRPS5	Mitochondrial ribosomal protein S5	2p11.2-q11.2	1.49E-02	-1.25
239809_at	---	---	---	4.45E-02	-1.25
37796_at	LRCH4/SAP25	leucine-rich repeats and calponin homology domain containing 4/ sin3A-binding protein 25	7q22 / 7q22.1	3.21E-03	-1.25
204491_at	PDE4D	Phosphodiesterase 4D, cAMP-specific	5q12	3.64E-02	-1.25
219624_at	BAG4	BCL2-associated athanogene 4	8p11.23	1.81E-03	-1.25

209356_x_at	EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	11q13.1	1.11E-02	-1.25
217521_at	---	---	---	4.36E-02	-1.25
202446_s_at	PLSCR1	Phospholipid scramblase 1	3q23	9.40E-03	-1.25
223393_s_at	TSHZ3	Teashirt zinc finger homeobox 3	19q12	3.90E-02	-1.25
230945_at	---	---	---	4.34E-04	-1.25
242949_x_at	CCDC157	Coiled-coil domain containing 157	22q12.2	8.04E-03	-1.26
213275_x_at	CTSB	Cathepsin B	8p22	1.09E-02	-1.26
209776_s_at	SLC19A1	Solute carrier family 19, member 1	21q22.3	3.25E-02	-1.26
234470_at	LOC143506	SSU72 RNA polymerase II CTD phosphatase homolog pseudogene	11p15.4	4.14E-03	-1.26
235536_at	SNORD89	Small nucleolar RNA, C/D box 89	2q11.2	2.45E-02	-1.26
1558887_at	ZNF321	Zinc finger protein 321	19q13.41	2.00E-02	-1.26
202534_x_at	DHFR	Dihydrofolate reductase	5q11.2-q13.2	4.14E-02	-1.26
241832_at	FAM98A	Family with sequence similarity 98, member A	2p22.3	4.14E-02	-1.26
1557322_at	ZNF221/ ZNF230	Zinc finger protein 221/zinc finger protein 230	19q13.2 / 19q13.31	2.63E-02	-1.26
205878_at	POU6F1	POU class 6 homeobox 1	12q13.13	1.02E-02	-1.26
63825_at	ABHD2	Abhydrolase domain containing 2	15q26.1	6.88E-03	-1.26
1560758_at	---	---	---	7.55E-03	-1.26
243433_at	---	---	---	1.75E-02	-1.27
209311_at	BCL2L2	BCL2-like 2	14q11.2-q12	4.85E-03	-1.27
1570048_at	DNAJC24	DnaJ homolog, subfamily C, member 24	11p13	3.61E-02	-1.27
202607_at	NDST1	N-deacetylase/N-sulfotransferase 1	5q33.1	2.21E-02	-1.27
235483_at	STX3	Syntaxin 3	11q12.1	1.56E-02	-1.27
1557788_a_at	C9orf130	Chromosome 9 open reading frame 130	9q22.32	3.69E-02	-1.27
219227_at	CCNJL	Cyclin J-like	5q33.3	3.18E-02	-1.27
242842_at	---	---	---	2.87E-03	-1.27
230094_at	---	---	---	7.64E-03	-1.27
205040_at	ORM1	Orosomucoid 1	9q31-q32	4.86E-02	-1.27
202572_s_at	DLGAP4	Discs, large homolog-associated protein 4	20q11.23	5.49E-03	-1.27

231732_at	SMPD3	Sphingomyelin phosphodiesterase 3, neutral membrane	16q22.1	6.20E-03	-1.27
217646_at	SURF1	Surfeit 1	9q34.2	3.10E-02	-1.27
206351_s_at	PEX10	Peroxisomal biogenesis factor 10	1p36.32	3.23E-02	-1.27
1555214_a_at	CLEC7A	C-type lectin domain family 7, member A	12p13.2	4.15E-02	-1.28
230827_at	---	---	---	3.75E-02	-1.28
1569257_at	FMNL1	Formin-like 1	17q21	4.45E-02	-1.28
209333_at	ULK1	Unc-51-like kinase 1	12q24.3	6.31E-03	-1.28
227080_at	ZNF697	Zinc finger protein 697	1p12	1.78E-02	-1.28
214014_at	CDC42EP2	CDC42 effector protein 2	11q13	4.53E-02	-1.28
240405_at	---	---	---	4.78E-02	-1.28
238446_at	GUSBP3	Glucuronidase, beta pseudogene 3	5q13.2	3.87E-02	-1.28
203765_at	GCA	Grancalcin, EF-hand calcium binding protein	2q24.2	4.35E-02	-1.28
215847_at	HERC2P3	Hect domain and RLD 2 pseudogene 3	15q11.1	2.82E-02	-1.28
228846_at	MXD1	MAX dimerization protein 1	2p13-p12	4.22E-02	-1.28
220034_at	IRAK3	Interleukin-1 receptor-associated kinase 3	12q14.3	3.85E-02	-1.28
212235_at	PLXND1	Plexin D1	3q22.1	2.31E-02	-1.28
217644_s_at	SOS2	Son of sevenless homolog 2	14q21	2.47E-02	-1.28
201905_s_at	CTDSPL	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	3p21.3	4.39E-03	-1.28
240240_at	---	---	---	3.42E-02	-1.28
240258_at	ENO1	Enolase 1	1p36.2	9.44E-03	-1.28
223553_s_at	DOK3	Docking protein 3	5q35.3	2.91E-02	-1.28
204564_at	PCGF3	Polycomb group ring finger 3	4p16.3	1.48E-02	-1.29
209681_at	SLC19A2	Solute carrier family 19, member 2	1q23.3	2.40E-02	-1.29
36829_at	PER1	Period homolog 1	17p13.1	2.80E-03	-1.29
221477_s_at	SOD2	Superoxide dismutase 2, mitochondrial	6q25.3	3.47E-02	-1.29
204099_at	SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3	7q35-q36	3.38E-02	-1.29
1561654_at	---	---	---	4.65E-02	-1.29

219583_s_at	SPATA7	Spermatogenesis associated 7	14q31.3	1.47E-02	-1.29
212666_at	SMURF1	SMAD specific E3 ubiquitin protein ligase 1	7q22.1	6.52E-03	-1.29
205145_s_at	MYL5	Myosin, light chain 5, regulatory	4p16.3	4.00E-02	-1.29
229130_at	---	---	---	4.91E-02	-1.29
205051_s_at	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	4q11-q12	2.58E-02	-1.29
239023_at	---	---	---	3.25E-02	-1.29
229373_at	---	---	---	2.44E-02	-1.29
223882_at	FAM172A	Family with sequence similarity 172, member A	5q15	4.03E-02	-1.29
235219_at	C5orf55	Chromosome 5 open reading frame 55	5p15.33	8.60E-03	-1.30
225262_at	FOSL2	FOS-like antigen 2	2p23.3	6.72E-03	-1.30
235685_at	---	---	---	9.13E-03	-1.30
201425_at	ALDH2	Aldehyde dehydrogenase 2 family	12q24.2	1.21E-02	-1.30
205241_at	SCO2	SCO cytochrome oxidase deficient homolog 2	22q13.33	3.67E-02	-1.30
202497_x_at	SLC2A3	Solute carrier family 2, member 3	12p13.3	2.74E-02	-1.30
1554067_at	C12orf66	Chromosome 12 open reading frame 66	12q14.2	2.20E-02	-1.30
220795_s_at	BEGAIN	Brain-enriched guanylate kinase-associated homolog	14q32.2	1.03E-02	-1.30
205708_s_at	TRPM2	Transient receptor potential cation channel, subfamily M, member 2	21q22.3	2.46E-02	-1.30
221042_s_at	CLMN	Calmin	14q32.13	1.95E-02	-1.30
215229_at	LOC100129973	Hypothetical LOC100129973	15q21.2	2.86E-02	-1.30
223916_s_at	BCOR	BCL6 corepressor	Xp11.4	2.51E-02	-1.30
225191_at	CIRBP	Cold inducible RNA binding protein	19p13.3	3.40E-02	-1.30
229966_at	EWSR1	Ewing sarcoma breakpoint region 1	22q12.2	9.59E-03	-1.30
213700_s_at	---	---	---	1.10E-02	-1.30
209204_at	LMO4	LIM domain only 4	1p22.3	1.02E-02	-1.30
204042_at	WASF3	WAS protein family, member 3	13q12	2.36E-02	-1.31
1552327_at	ARMCX4	Armadillo repeat containing, X-linked 4	Xq22.1	6.31E-03	-1.31
239373_at	---	---	---	4.64E-02	-1.31
211097_s_at	PBX2	Pre-B-cell leukemia homeobox 2	6p21.3	5.09E-03	-1.31

222088_s_at	SLC2A14/ SLC2A3	Solute carrier family 2 (SLC2), member 14/ SLC2, member 3	12p13.3 / 12p13.31	1.67E-02	-1.31
222796_at	PTCD1	Pentatricopeptide repeat domain 1	7q22.1	1.05E-02	-1.31
213274_s_at	CTSB	Cathepsin B	8p22	1.83E-02	-1.31
215760_s_at	SBNO2	Strawberry notch homolog 2	19p13.3	3.45E-02	-1.31
1557036_at	ZBTB1	Zinc finger and BTB domain containing 1	14q23.3	3.77E-02	-1.31
206277_at	P2RY2	Purinergic receptor P2Y, G-protein coupled, 2	11q13.5-q14.1	3.52E-02	-1.31
207387_s_at	GK	Glycerol kinase	Xp21.3	1.26E-02	-1.32
219761_at	CLEC1A	C-type lectin domain family 1, member A	12p13.2	1.75E-02	-1.32
217739_s_at	NAMPT	Nicotinamide phosphoribosyltransferase	7q22.3	2.19E-02	-1.32
244267_at	---	---	---	4.20E-02	-1.32
205119_s_at	FPR1	Formyl peptide receptor 1	19q13.4	2.01E-02	-1.32
217817_at	ARPC4	Actin related protein 2/3 complex, subunit 4, 20kDa	3p25.3	2.72E-02	-1.32
202498_s_at	SLC2A3	Solute carrier family 2, member 3	12p13.3	3.61E-02	-1.32
1569500_at	---	---	---	4.03E-02	-1.32
215850_s_at	NDUFA5	NADH dehydrogenase 1 alpha subcomplex, 5, 13kDa	7q32	1.53E-02	-1.32
240451_at	---	---	---	3.46E-02	-1.32
244025_at	---	---	---	3.10E-02	-1.33
232311_at	B2M	Beta-2-microglobulin	15q21-q22.2	2.47E-02	-1.33
222026_at	RBM3	RNA binding motif protein 3	Xp11.2	6.79E-03	-1.33
209185_s_at	IRS2	Insulin receptor substrate 2	13q34	2.15E-02	-1.33
240569_at	---	---	---	4.14E-02	-1.33
232837_at	KIF13A	Kinesin family member 13A	6p23	1.97E-02	-1.33
238635_at	C5orf28	Chromosome 5 open reading frame 28	5p12	4.94E-02	-1.33
228933_at	NHS	Nance-Horan syndrome	Xp22.13	4.09E-02	-1.33
210139_s_at	PMP22	Peripheral myelin protein 22	17p12	3.24E-03	-1.33
1560014_s_at	PDXDC1	Pyridoxal-dependent decarboxylase domain containing 1	16p13.11	3.09E-02	-1.33
226880_at	---	---	---	1.06E-02	-1.34
229733_s_at	---	---	---	5.28E-04	-1.34

231323_at	PSMB2	Proteasome subunit, beta type, 2	1p34.2	2.65E-02	-1.34
239313_at	LOC401320	Hypothetical LOC401320	7p15.1	4.79E-02	-1.34
224444_s_at	C1orf97	Chromosome 1 open reading frame 97	1q32.3	3.44E-03	-1.34
235678_at	GM2A	GM2 ganglioside activator	5q33.1	3.59E-02	-1.34
231777_at	CSNK2B/LY6G5B	Casein kinase 2, beta polypeptide/ lymphocyte antigen 6 complex, locus G5B	6p21-p12|6p21.3 / 6p21.3	2.29E-02	-1.34
226076_s_at	MBD6	Methyl-CpG binding domain protein 6	---	1.62E-02	-1.34
1559094_at	FBXO9	F-box protein 9	6p12.3-p11.2	2.71E-02	-1.35
242557_at	NCRNA00171	Non-protein coding RNA 171	6p22.1	4.01E-02	-1.35
223677_at	ATG10	ATG10 autophagy related 10 homolog	5q14.1	2.90E-02	-1.35
205568_at	AQP9	Aquaporin 9	15q	2.76E-02	-1.35
231260_at	LOC386758	Hypothetical LOC386758	19q13.43	3.38E-03	-1.35
235749_at	UGGT2	UDP-glucose glycoprotein glucosyltransferase 2	13q32.1	3.84E-02	-1.36
213265_at	PGA3/ PGA4/ PGA5	Pepsinogen 3, group I/ pepsinogen 4, group I / pepsinogen 5, group I	11q12.2 / 11q13	3.64E-02	-1.36
217208_s_at	DLG1	Discs, large homolog 1	3q29	1.21E-02	-1.36
229520_s_at	C14orf118	Chromosome 14 open reading frame 118	14q22.1-q24.3	1.84E-02	-1.36
236470_at	---	---	---	4.97E-03	-1.36
244046_at	URGCP	Upregulator of cell proliferation	7p13	2.13E-02	-1.36
230256_at	C1orf104	Chromosome 1 open reading frame 104	1q22	4.02E-02	-1.36
232284_at	PSMD6	Proteasome 26S subunit, non-ATPase, 6	3p14.1	1.40E-02	-1.36
241233_x_at	C21orf81	Ankyrin repeat domain 20 family, member A3 pseudogene	21q11.2	3.73E-02	-1.37
235263_at	STAG3L1	Stromal antigen 3-like 1	7q11.23	6.57E-03	-1.37
236386_at	LOC100506501	Hypothetical LOC100506501	---	8.39E-04	-1.37
217257_at	SH3BP2	SH3-domain binding protein 2	4p16.3	1.08E-02	-1.37
201883_s_at	B4GALT1	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	9p13	2.62E-02	-1.37
207535_s_at	NFKB2	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2	10q24	2.56E-02	-1.38
244484_at	---	---	---	1.03E-03	-1.38
243495_s_at	---	---	---	4.66E-02	-1.38
1556590_s_at	---	---	---	3.74E-02	-1.38

237208_at	WDR61	WD repeat domain 61	15q25.1	8.90E-03	-1.38
236487_at	SCLT1	Sodium channel and clathrin linker 1	4q28.2	1.49E-02	-1.38
213121_at	SNRNP70	Small nuclear ribonucleoprotein 70kDa	19q13.3	3.26E-03	-1.38
240392_at	---	---	---	3.42E-02	-1.39
238376_at	---	---	---	4.28E-02	-1.39
236985_at	---	---	---	4.95E-02	-1.39
204647_at	HOMER3	Homer homolog 3	19p13.11	3.15E-02	-1.39
244677_at	---	---	---	1.87E-03	-1.40
214657_s_at	NEAT1	Nuclear paraspeckle assembly transcript 1	11q13.1	4.20E-02	-1.40
218546_at	C1orf115	Chromosome 1 open reading frame 115	1q41	1.87E-03	-1.40
241938_at	QKI	Quaking homolog, KH domain RNA binding	6q26-q27	2.52E-02	-1.40
233217_at	---	---	---	4.98E-04	-1.40
202861_at	PER1	Period homolog 1	17p13.1	1.30E-02	-1.40
219859_at	CLEC4E	C-type lectin domain family 4, member E	12p13.31	4.14E-02	-1.40
223608_at	EFCAB2	EF-hand calcium binding domain 2	1q44	2.91E-02	-1.40
1568780_at	---	---	---	4.59E-02	-1.41
215761_at	DMXL2	Dmx-like 2	15q21.2	3.41E-02	-1.41
202499_s_at	SLC2A3	Solute carrier family 2, member 3	12p13.3	3.09E-03	-1.41
207993_s_at	CHP	Calcium binding protein P22	15q13.3	1.14E-02	-1.41
209298_s_at	ITSN1	Intersectin 1	21q22.1-q22.2	2.58E-02	-1.41
201896_s_at	PSRC1	Proline/serine-rich coiled-coil 1	1p13.3	4.92E-03	-1.42
1557100_s_at	HECTD1	HECT domain containing 1	14q12	4.73E-03	-1.42
211688_x_at	KIR3DL1/ KIR3DL2/ LOC727787	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail (KIR3DL), 1/ KIR3DL, 2 /KIR3DL, 2-like	19q13.4	4.79E-02	-1.42
209446_s_at	C7orf44	Chromosome 7 open reading frame 44	7p13	2.06E-02	-1.42
241631_at	FLJ10357	Protein SOLO	14q11.2	2.03E-02	-1.42
241627_x_at	ARHGEF40	Rho guanine nucleotide exchange factor 40	14q11.2	4.59E-02	-1.43
219837_s_at	CYTL1	Cytokine-like 1	4p16-p15	2.48E-02	-1.44
240573_at	LOC374443	CLR pseudogene	12p13.31	3.22E-02	-1.44

1559018_at	PTPRE	Protein tyrosine phosphatase, receptor type, E	10q26	2.46E-02	-1.44
243469_at	---	---	---	4.58E-02	-1.44
225283_at	ARRDC4	Arrestin domain containing 4	15q26.3	1.24E-02	-1.44
215188_at	STK24	Serine/threonine kinase 24	13q31.2-q32.3	2.29E-02	-1.44
223940_x_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	11q13.1	4.09E-02	-1.45
209263_x_at	TSPAN4	Tetraspanin 4	11p15.5	4.21E-03	-1.45
228030_at	RBM6	RNA binding motif protein 6	3p21.3	1.29E-02	-1.46
208426_x_at	KIR2DL4/ KIR2DL5A/ LOC100287534	killer cell immunoglobulin-like receptor (KIR), two domains, long cytoplasmic tail (KIR2DL), 4/ KIR2DL, 5A/ KIR2DL, 4-like	19p13.3 / 19q13.4	2.50E-02	-1.46
227961_at	CTSB	Cathepsin B	8p22	2.10E-03	-1.46
227762_at	---	---	---	3.69E-02	-1.46
1560013_at	PDXDC1	Pyridoxal-dependent decarboxylase domain containing 1	16p13.11	2.88E-03	-1.46
230058_at	SDCCAG3	Serologically defined colon cancer antigen 3	9q34.3	5.13E-03	-1.47
222303_at	---	---	---	2.73E-02	-1.47
235798_at	TMEM170B	Transmembrane protein 170B	6p24.2	4.18E-02	-1.48
216907_x_at	KIR3DL1/ KIR3DL2/ LOC727787	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail (KIR3DL), 1/ KIR3DL, 2 /KIR3DL, 2-like	19q13.4	4.12E-02	-1.48
212574_x_at	C19orf6	Chromosome 19 open reading frame 6	19p13.3	1.39E-02	-1.49
203634_s_at	CPT1A	Carnitine palmitoyltransferase 1A	11q13.2	3.14E-02	-1.49
221541_at	CRISPLD2	Cysteine-rich secretory protein LCCL domain containing 2	16q24.1	1.04E-02	-1.49
219570_at	KIF16B	Kinesin family member 16B	20p11.23	1.53E-02	-1.50
205041_s_at	ORM1/ORM2	Orosomucoid 1/orosomucoid 2	9q31-q32 / 9q32	2.05E-02	-1.50
1554624_a_at	SIRPB1	Signal-regulatory protein beta 1	20p13	3.16E-02	-1.50
229307_at	---	---	---	1.69E-02	-1.52
228285_at	TDRD9	Tudor domain containing 9	14q32.33	4.74E-02	-1.52
207314_x_at	KIR3DL1/ KIR3DL2/ LOC727787	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail (KIR3DL), 1/ KIR3DL, 2 /KIR3DL, 2-like	19q13.4	2.45E-02	-1.53
219359_at	ATHL1	ATH1, acid trehalase-like 1	11p15.5	2.40E-02	-1.53
243395_at	---	---	---	3.73E-03	-1.53
204908_s_at	BCL3	B-cell CLL/lymphoma 3	19q13.1-q13.2	1.24E-02	-1.53

234989_at	---	---	---	1.97E-02	-1.53
206237_s_at	NRG1	Neuregulin 1	8p12	3.98E-02	-1.53
223578_x_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	11q13.1	7.49E-03	-1.53
235157_at	---	---	---	2.14E-02	-1.54
216905_s_at	ST14	Suppression of tumorigenicity 14	11q24-q25	3.77E-03	-1.55
224568_x_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	11q13.1	1.98E-02	-1.56
227510_x_at	MALAT1	Metastasis associated lung adenocarcinoma transcript 1	11q13.1	1.12E-02	-1.57
204036_at	LPAR1	Lysophosphatidic acid receptor 1	9q31.3	3.45E-02	-1.57
202464_s_at	PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	10p15.1	2.35E-03	-1.58
1569607_s_at	ANKRD20A1/ ANKRD20A2/ ANKRD20A3A4/ C21orf81/ LOC100132733/ LOC284232	Ankyrin repeat domain 20 family (ANKRD20) member A1/ ANKRD20 member A2/ ANKRD20 member A3/ ANKRD20 member A4/ ANKRD20 member A3 pseudogene/ ANKRD20 member A3-like/ ANKD20 member A2 pseudogene	13q11 / 1q21.1 / 21q11.2 / 9p12 / 9q13 / 9q21.11	2.21E-02	-1.61
213986_s_at	C19orf6	Chromosome 19 open reading frame 6	19p13.3	5.46E-03	-1.62
1569599_at	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1	21q11	3.84E-02	-1.63
209264_s_at	TSPAN4	Tetraspanin 4	11p15.5	1.30E-02	-1.65
1558796_a_at	---	---	---	4.16E-02	-1.65
236244_at	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	1q44	2.42E-02	-1.67
232530_at	PLD1	Phospholipase D1, phosphatidylcholine-specific	3q26	1.06E-02	-1.72
236495_at	---	---	---	2.87E-03	-1.72
206343_s_at	NRG1	Neuregulin 1	8p12	3.12E-02	-1.76
227062_at	---	---	---	5.81E-03	-1.77
205239_at	AREG	Amphiregulin	4q13-q21	4.90E-02	-1.80
243296_at	NAMPT	Nicotinamide phosphoribosyltransferase	7q22.3	5.24E-04	-1.86
1558795_at	---	---	---	1.95E-02	-2.16
209686_at	S100B	S100 calcium binding protein B	21q22.3	7.92E-03	-2.25
205767_at	EREG	Epiregulin	4q13.3	2.99E-02	-2.39

Supplementary Table 3. Significant canonical pathways found enriched in the IPA analysis. Canonical pathways with $P \leq 0.05$ are shown. Genes that were among the top 30 most differentially expressed between cases and controls (Table 2) are highlighted in bold. The ratio corresponds to the number of molecules differentially expressed in our dataset divided by the total number of molecules in that pathway.

Ingenuity canonical pathway	Molecules	*P*	Ratio
Lipid antigen presentation by CD1	B2M, CALR, **TRD@**, CD1C	4.34E-04	4/22
RhoA signalling	**LPAR1**, EZR, MYL5, ARPC4, MYL4, PPP1CB, CDC42EP2, TTN, **PLD1**	1.72E-03	9/22
Cdc42 signalling	B2M, **TRD@**, MYL5, ARPC4, **HLA-DQA1**, MYL4, PPP1CB, CDC42EP2, EXOC3	3.51E-03	9/143
Role of JAK2 in hormone-like cytokine signalling	SOCS6, PTPN1, IRS2, SOCS7	8.17E-03	4/34
Caveolar-mediated endocytosis signalling	B2M, CD55, ITSN1, PTPN1, CAV1, COPE	8.80E-03	6/81
Antigen presentation pathway	B2M, CALR, **HLA-DQA1**, TAPBP	9.12E-03	4/40
Neuregulin signalling	ERBB4, DCN, SOS2, **NRG1**, **EREG**, **AREG/AREGB**	2.04E-02	6/95
Primary immunodeficiency signalling	RFXAP, IGHE, **IGHM**, AICDA	2.24E-02	4/55
Phospholipase C signalling	**TRD@**, MYL5, GPLD1, SOS2, CHP, MYL4, PPP1CB, RAPGEF3, NFKB2, **PLD1**, GNAL	2.29E-02	11/244
Fcγ receptor-mediated phagocytosis in macrophages and monocytes	GAB2, EZR, GPLD1, ARPC4, FGR, **PLD1**	2.75E-02	6/94
Nur77 signalling in T lymphocytes	**TRD@**, CHP, **HLA-DQA1**, SIN3B	2.82E-02	4/57
OX40 signalling pathway	B2M, **TRD@**, **HLA-DQA1**, NFKB2	2.82E-02	4/62
Cellular effects of sildenafil (Viagra)	KCNN4, MYL5, MYL4, PPP1CB, PDE4D, GNAL, CACNA1A	4.68E-02	7/132
Aryl hydrocarbon receptor signalling	CCNE2, GSTM3, DHFR, NFKB2, ALDH6A1, HSPB1, CHEK1	4.79E-02	7/141

Supplementary Table 4. Characterization of the investigated SNPs. The names, chromosome (chr.) and base pair (bp) positions, the Hardy Weinberg equilibrium (HWE) P in the control group, and case-control minor allele frequencies (MAF) in the Iranian dataset are indicated in this table for the genotyped SNPs.

Gene/locus	SNP ID	SNP	Chr.	Position (bp)	Alleles	HWE *P*	MAF in controls	MAF in cases
EREG-AREG	1	rs7687621	4	75249826	C:T	1.000	0.239	0.225
	2	rs2061509	4	75250146	C:T	0.992	0.239	0.227
	3	rs6816797	4	75258678	A:G	0.350	0.221	0.209
	4	rs6845297	4	75259405	G:A	0.664	0.423	0.461
	5	rs12641287	4	75259691	A:G	0.882	0.343	0.329
	6	rs9992496	4	75263673	G:A	0.029	0.354	0.372
	7	rs1494882	4	75268407	T:C	0.560	0.388	0.377
	8	rs7693713	4	75272863	C:T	0.232	0.179	0.176
	9	rs12503501	4	75299835	G:C	0.942	0.495	0.513
	10	rs1691273	4	75323645	T:C	0.307	0.367	0.372
	11	rs1691274	4	75324290	T:C	0.139	0.385	0.375
	12	rs4350980	4	75326268	G:C	0.478	0.122	0.127
NRG1	13	rs4489285	8	31398121	A:G	0.017	0.302	0.339
	14	rs1557800	8	31434385	C:T	0.238	0.113	0.125
	15	rs10954811	8	31586453	G:A	0.199	0.179	0.196
	16	rs2345991	8	31608795	C:T	0.755	0.382	0.369
	17	rs1503486	8	31609577	A:G	0.623	0.440	0.438
	18	rs4733272	8	31612816	A:G	0.398	0.355	0.335
	19	rs16878317	8	31657854	C:T	0.348	0.157	0.156
	20	rs776385	8	31736625	C:T	0.862	0.291	0.302
	21	rs383632	8	31806048	C:T	0.368	0.180	0.153
	22	rs1462891	8	31830933	C:T	0.739	0.261	0.231
	23	rs17683983	8	31846839	C:A	0.082	0.379	0.395
	24	rs1462875	8	31860194	C:A	0.171	0.368	0.348
	25	rs956203	8	31910758	A:G	0.599	0.366	0.364
	26	rs16878764	8	31936552	T:G	0.341	0.369	0.368
	27	rs10503899	8	31947234	A:G	0.017	0.411	0.424

Supplementary Table 5. Association results. For each SNP tested, the crude allelic P-value (Pa) is indicated. Significant P-values (≤0.05) are highlighted in bold, and the respective odds ratios (OR) and 95% confidence intervals (CI) are shown.

Gene/locus	SNP ID	SNP	P_a	Allele: OR [95% CI]
EREG-AREG	1	rs7687621	3.36E-01	NS
	2	rs2061509	4.12E-01	NS
	3	rs6816797	3.86E-01	NS
	4	rs6845297	**2.51E-02**	A: 0.86 [0.75-0.98]
	5	rs12641287	4.02E-01	NS
	6	rs9992496	2.66E-01	NS
	7	rs1494882	4.93E-01	NS
	8	rs7693713	8.04E-01	NS
	9	rs12503501	2.84E-01	NS
	10	rs1691273	7.51E-01	NS
	11	rs1691274	5.65E-01	NS
	12	rs4350980	6.61E-01	NS
NRG1	13	rs4489285	**2.01E-02**	G: 0.85 [0.73-0.97]
	14	rs1557800	2.95E-01	NS
	15	rs10954811	1.95E-01	NS
	16	rs2345991	4.25E-01	NS
	17	rs1503486	9.15E-01	NS
	18	rs4733272	2.07E-01	NS
	19	rs16878317	9.66E-01	NS
	20	rs776385	4.74E-01	NS
	21	rs383632	**2.78E-02**	T: 1.22 [1.02-1.46]
	22	rs1462891	**3.92E-02**	T: 1.18 [1.01-1.37]
	23	rs17683983	3.54E-01	NS
	24	rs1462875	2.10E-01	NS
	25	rs956203	8.98E-01	NS
	26	rs16878764	9.47E-01	NS
	27	rs10503899	4.19E-01	NS

Supplementary Table 6. Gene-gene interaction models obtained using the multifactor-dimensionality reduction (MDR) method in BD susceptibility. The testing balanced accuracy (TBA) indicates the percentage of individuals correctly classified as BD or controls, and the cross validation consistency (CVC) indicates the number of times (out of 10 cross-validation subsets) that the model was selected. Ppermut is the P-value calculated after 1,000 permutations.

Number of loci	Best model	CVC	TBA	P_{permut}
2	rs6845297 (downstream *EREG*), rs1462875 (*NRG1*)	5/10	0.536	1.61E-01
3	rs6845297 (downstream *EREG*), rs2345991 (*NRG1*), rs1462875 (*NRG1*)	6/10	0.545	7.75E-02
4	rs6845297 (downstream *EREG*), rs9992496 (downstream *EREG*), rs2345991 (*NRG1*), rs956203 (*NRG1*)	5/10	0.566	8.50E-03

CHAPTER 6

FUT2: filling the gap between genes and environment in Behçet's disease?

(Xavier *et al.*, Manuscript under preparation)

FUT2: filling the gap between genes and environment in Behçet's disease?

Joana M. Xavier,[1,2] Inês Sousa[1,2], Mafalda Matos[1,2], João Sobral[1,2], Fahmida Ghaderibarim,[3] Farhad Shahram,[3] Abdolhadi Nadji,[3] Manuela Oliveira[4], Niloofar Mojarad Shafiee,[3] Bahar Sadeghi Abdollahi,[3] Fereydoun Davatchi,[3] and Sofia A. Oliveira[1,2,*]

Affiliations

[1]Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal.

[2]Instituto Gulbenkian de Ciência, Oeiras, Portugal.

[3]Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran.

[4]Universidade de Évora, Portugal.

Key-words

Behçet's disease, *FUT2*, GWAS, DNA pooling, ABO secretor genotype

6.1 ABSTRACT

Background Behçet's disease (BD) is a multisystemic immune and inflammatory disorder whose aetiology remains unclear. In order to identify novel susceptibility loci, we performed the first genome-wide association study (GWAS) for BD in the Iranian population using a DNA pooling strategy.

Methods Two replicate pools of 292 BD cases and of 294 age- and sex-matched controls were allelotyped in quadruplicate on the Affymetrix Human SNP Array 6.0 arrays. The SNPs were ranked by relative allele score difference between cases and controls and 47 out of the 51 top markers were technically validated through individually genotyping. Replication of validated SNPs was performed in an independent Iranian dataset of 684 cases and 532 controls.

Results In addition to the *HLA-B* locus, rs7528842 in a gene desert on chromosome 1p21.2, and rs632111 at the 3´UTR of *FUT2* were associated in both the discovery and replication datasets (individually and in combination). However, only the *FUT2* SNP has been associated in a previous GWAS for BD in Turkish. Fine-mapping of *FUT2* in the full Iranian dataset revealed additional associations in 5 coding SNPs (2.97E-06<*P*<1.34E-04), including the rs601338 non-sense (W143X) variant which, in Caucasians, determines the secretion of the H antigen (precursor of the ABO blood group antigens) in body fluids and on the intestinal mucosa. Meta-analysis with the published Turkish GWAS data strengthened the *FUT2* associations (4.78E-09<*P*<1.66E-07).

Discussion The non-secretor phenotype affects mucosal glycosylation, which may explain its known association with dysbiosis and altered susceptibility to infections. This different antigenic stimulation in early life and consequent increased propensity for auto-immunity and inflammation may contribute to BD development. While confirming the well-established association of the *HLA-B* locus with BD, this study established for the first time a putative link between genes and environment in the etiopathogenesis of BD.

6.2 INTRODUCTION

Behçet's disease is chronic multisystemic vasculitis whose main symptoms are recurrent orogenital ulcerations, ocular and skin lesions. BD is a complex disorder with genetic and environmental risk factors implicated in its aetiology, but its pathogenic mechanisms are still poorly understood [Kaya, 2004]. GWAS represent the gold-standard approach to study the genetic basis of complex traits and several GWAS have been published for BD. These studies identified novel associations with BD at *UBAC2, HLA-A, IL10, IL23R-IL12RB2, STAT4* and *GIMAP* loci [Fei *et al*., 2009, Remmers *et al*., 2010, Mizuki *et al*., 2010, Hou *et al*., 2012, Lee *et al*., 2013], which were replicated in different populations. However, these reports also showed little overlap between the top findings, highlighting the complexity of BD genetics and the importance of further research on the genetic underpinnings of this disease.

Since genotyping individually thousands of polymorphisms in the large datasets required for the study of complex traits is time-consuming, labour-intensive, and cost-prohibitive, we used a validated strategy that combines DNA pooling and microarray genotyping [Craig *et al*., 2005, Pearson *et al*., 2007; Melquist *et al*., 2007; Steer *et al*., 2007] to perform the first GWAS in the Iranian population. In this pooling approach, no individual genotypes are obtained but instead allelotyping is performed on pools of DNA from cases and controls. The power of this technique stems from the fact that the estimated allele frequencies from allelotyping are a good proxy for the true allelic frequencies [Abraham *et al*., 2008; Bossé *et al*., 2009; Bostrom *et al*., 2010]. Using different high-density genotyping platforms and numerous analysis methods to rank the polymorphisms, this pooling GWAS strategy has been validated by the replication of known associations while identifying new loci for other complex disorders [Abraham *et al*., 2008; Ricci *et al*., 2011; Gaj *et al*., 2012]. We hereby apply this approach in order to identify novel genes involved in BD susceptibility.

6.3 PATIENTS AND METHODS

6.3.1 Study subjects

The discovery dataset used in the GWAS includes 292 BD cases and 294 age- and gender-matched healthy controls (Table 1). The replication dataset is composed of 684 BD cases and 532 unrelated healthy controls also from the Iranian population. The main demographic and clinical characteristics of these samples are summarized in Table 1.

The participants were enrolled in the Behçet's Disease Clinic, Behçet's Unit, Rheumatology Research Center (Shariati Hospital) at the Tehran University of Medical Sciences, Tehran, Iran. This study received ethics approval from the ethics committee at the Tehran University for Medical Sciences, Iran.

All participants were informed of the study, provided informed written consent and the study was conducted according to the Declaration of Helsinki. Patients with age at onset of BD after 60 years were excluded. Controls were evaluated using the same evaluation procedures as the cases and selected when negative for BD, any other rheumatologic or autoimmune disorder, and oral aphthosis. The clinical and demographic features of the study participants were obtained by medical interview at the time of blood sampling and inspection of medical records. BD cases were selected as consecutive patients and the diagnosis of BD was made according to the revised International Criteria for Behçet's Disease [International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD), 2006].

6.3.2 Construction of DNA pools

Genomic DNA was extracted from whole blood samples using a salting out procedure and diluted in TE buffer. The concentrations of extracted DNA were determined by Nanodrop using absorbance readings at 260nm. DNA samples from 300 BD cases and 300 healthy control samples were diluted with sterile water to an approximate concentration of 85ng/ul. The concentration of each diluted sample was then determined in triplicate using the PicoGreen dsDNA Quantitation Reagent (Molecular Probes, Inc., Eugene, Oregon, USA) in a PerkinElmer top Fluoroscence reader (PerkinElmer, Inc., Waltham, USA). Samples with less than 60ng/ul were concentrated and quantified again in triplicate. Samples with greater than 3% of the sample standard deviation between replicates or with greater than two

standard deviations from the median volume to be pooled for each sample (8 BD cases and 6 healthy controls), were not included in the respective pool. 200ng of DNA from each sample that passed quality control (292 BD cases and 294 controls) was then added to either a case or a control pool. To minimize the errors associated with pipetting no less than 1.8µl of each sample was added to a pool. This procedure was repeated twice for each sample so as to prepare two replicates pools of cases and of controls.

Table 1. General characteristics of the Iranian discovery and replication datasets.

Characteristic	Discovery dataset		Replication dataset	
	Cases	Controls	Cases	Controls
N	292	294	684	532
Gender (n/N, % males)	176/292 (60.3)	175/294 (59.5)	336/684 (49.3)	175/532 (32.9)
Mean age at examination (years ± SD)	39.1 ± 3.8	39.7 ± 5.4	39.1 ± 12.8	40.9 ± 14.3
Mean age at diagnosis (years ± SD)	32.3 ± 5.7	-	30.0 ± 10.2	-
Oral aphthosis (n/N, %)	292/292 (100)	0/294 (0)	673/684 (98.4)	0/532 (0)
Genital aphthosis (n/N, %)	213/292 (72.9)	0/294 (0)	399/684 (58.3)	0/532 (0)
Skin lesions (n/N, %)	207/292 (70.9)	-	335/684 (49.0)	-
Ophthalmologic manifestations (n/N, %)	171/292 (58.6)	-	413/684 (60.6)	-
Joint manifestations (n/N, %)	86/291 (29.5)	-	216/684 (31.7)	-
Neurological manifestations (n/N, %)	17/292 (5.8)	-	43/684 (6.3)	-
Vascular involvement (n/N, %)	15/292 (5.1)	-	35/684 (5.1)	-
Pathergy phenomenon (n/N, %)	166/287 (57.8)	-	276/676 (40.8)	-
Family history of BD (n/N, %)	23/281 (8.2)	-	62/658 (9.4)	-

6.3.3 Genome-wide allelotyping with the Affymetrix platform

Genome-wide genotyping of over 906600 SNPs was performed using the Affymetrix Human SNP Arrays 6.0 (Santa Clara, California, USA). The arrays were processed at the Instituto Gulbenkian de Ciência's Affymetrix Core Facility following manufacturer's protocols. Because of the expected inter-array variation, four replicate arrays were allelotyped for each of the four pools. Probe intensity data was read from CEL files and Relative Allele Scores (RAS) were calculated for each quartet using the SNPMaP package [Davis *et al.*, 2009]. RAS corresponds to the ratio of the A probe to the sum of the A and B probes (where A is the major allele and B is the minor allele) and provides a quantitative index correlated to the allele frequency in pooled DNA. All copy number variation (CNV)-related SNPs and all the polymorphisms on sex chromosomes were excluded from further analyses.

A Pearson's correlation coefficient was calculated between the average of the RAS values of cases and controls using the R freeware (http://cran.r-project.org/).

6.3.4 Individual genotyping

56 of the top SNPs with more than 8.5% of $|RAS_{diff}|$ and minor allele frequency (MAF) above 5% in the CEU HapMap population (http://www.hapmap.org/, release 27) were selected for technical validation through individual genotyping of the GWAS samples. Additionally, 5 haplotype tagging SNPs (htSNPs) located at *FUT2* together with the remaining coding SNPs were selected for fine-mapping follow-up. htSNPs in *FUT2* (chr.19: 53,891kb to 53,902kb) were identified in Haploview 4.2 [Barrett *et al*., 2005] using genotypes of 30 European (CEU) family trios (v3, release 27) and with the following options: pairwise mode, $r^2>0.95$ and minimum MAF of 0.05.

Individual samples were genotyped using Sequenom's (San Diego, USA) iPlex assay (primer extension of multiplex products with detection by matrix assisted laser desorption/ionization time-of-flight mass spectrometry) following manufacturer's protocol and detected in a Sequenom MassArray K2 platform. The primer sequences (Supplementary Table 1) were designed using Sequenom's MassARRAY ® Assay Design 3.0 software. All the genotyping was performed in the Genomics Unit of the Instituto Gulbenkian de Ciência. Two SNPs from the GWAS technical validation (rs1355282 and rs12364790) failed in the assay design phase and were not genotyped.

Extensive quality control was performed using eight HapMap (http://hapmap.ncbi.nlm.nih.gov/) controls of diverse ethnic affiliation, sample duplication within and across plates, Mendelian inheritance check in three large pedigrees, Hardy-Weinberg equilibrium (HWE) in the control group ($P>0.001$), and a minimum of 90% call rate for each SNP. Genotype determinations were performed blinded to affection status. Individual DNA samples with genotyping success rate across all SNPs < 85% were also excluded. Three SNPs from the GWAS technical validation (rs4954974, rs7577271, rs1592050) and one SNP from *FUT2* fine-mapping (rs485186) were excluded due to quality control issues (e.g., genotyping errors, low call rate).

6.3.5 Association analyses

Unpaired Student's t-tests and $\chi 2$ tests were used to compare quantitative and qualitative clinical and demographic data, respectively, between BD cases and controls. The adapted Manhattan plot was created using the R freeware. $\chi 2$ tests for HWE in controls and LD plots were performed using Haploview 4.2. The association of mitochondrial SNPs with

BD risk was tested using a Pearson's $\chi 2$ test. Association analyses unadjusted and adjusted for gender and pairwise conditional analysis were performed with logistic regression using the SNPassoc v.1.4-9 package [González *et al*., 2007] implemented in R. Significantly associated SNPs ($P \leq 0.05$) in the technical validation were considered eligible for further analyses. No correction for multiple testing was applied.

A Pearson's correlation coefficient between the average RAS values derived from the pooling data and the allele frequencies obtained by individual genotyping in the discovery dataset was calculated in R.

A combined analysis with the two Iranian datasets was performed for the SNPs associated in the GWAS, after correction for multiple testing, and for the *FUT2* fine-mapping SNPs. The GWAS association results (effective genotyping and imputation results) for *FUT2* in the Turkish dataset [Remmers *et al*., 2010, Kirino *et al*., 2013] were obtained from Dr. Elaine F. Remmers by personal communication.

6.3.6 Meta-analyses

Fixed effects (Mantel-Haenszel) meta-analyses were performed using the rmeta package in R.2.7.2 and PLINK v1.07 [Purcell *et al*., 2007]. Heterogeneity between studies was measured by the $\chi 2$-based Cochran's Q statistic (heterogeneity was considered significant for $P_{het} < 0.01$).

6.4 RESULTS

6.4.1 DNA pooling and GWAS

The general characteristics of the discovery dataset used in the GWAS are shown in Table 1. DNA samples from 294 controls and 292 age- and sex-matched cases met our quality controls and were therefore pooled in equimolar amounts in duplicate and allelotyped in quadruplicate on Affymetrix Human SNP Arrays 6.0 assaying over 906600 SNPs. We chose to make two replicate pools of cases and controls and hybridise each of them four times (total of 16 arrays) rather than constructing several smaller pools and hybridising them on single arrays, as this method has been shown to be more powerful [Macgregor *et al*., 2008]. The 16 arrays exhibited a similar RAS distribution analysed using box-and-whisker plot (data not shown).

A total of 868,254 SNPs which are not in copy number variants or in sex chromosomes were further analysed. The average of the RAS values over the 8 case and 8 control arrays showed a strong Pearson correlation with each other (r=0.996), suggesting a low technical variability of the method. Figure 1 depicts the absolute value of the RAS difference ($|RAS_{diff}|$) between cases and controls for all the assayed genetic markers. As expected, the SNPs with largest $|RAS_{diff}|$ are located in the major histocompatibility complex (MHC) region. Additionally, there are 55 autosomal SNPs located outside of chromosome 6p21.33 and 1 mitochondrial marker with a $|RAS_{diff}| > 8.5\%$ (Supplementary Table 2).

To validate the pool construction, the top 3 SNPs in the MHC locus and the 53 SNPs with a $|RAS_{diff}|>8.5\%$ and a MAF>5% in the CEU HapMap population were selected for technical validation by individual genotyping in the discovery dataset. All of these SNPs had a RAS variance between replicates lower than 0.25%. Five SNPs failed assay design or quality control, but 47 out of the 51 SNPs successfully genotyped (76%) were technically validated since they were associated with BD at the conventional *P*-value of 0.05 (Supplementary Table 3). For these 51 SNPs, the RAS values showed a Pearson's correlation of 0.853 with the allele frequencies determined by individual genotyping.

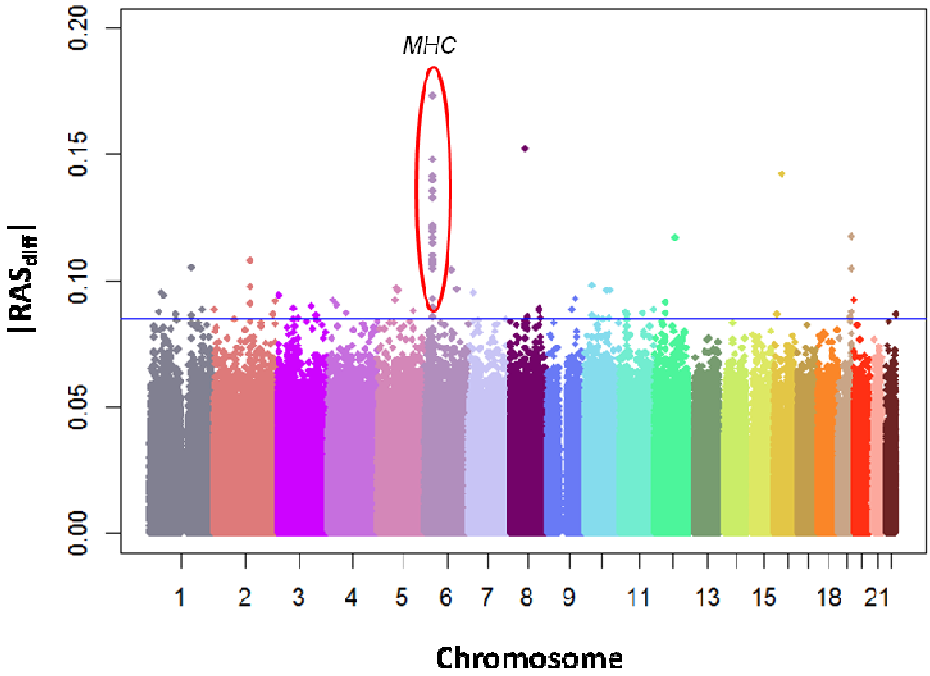


Figure 1. Plot of $|RAS_{diff}|$ against chromosomal location for the BD GWAS in pools. The absolute value of the relative allele score (RAS) difference between cases and controls is shown for 868,147 autossomal SNPs allelotyped in 292 Iranian cases and 294 controls, ordered by chromosomal position. The blue line is drawn at $|RAS_{diff}|$=0.085.

6.4.2 Independent replication of GWAS associated SNPs

For the SNPs that were technically validated, the next step is to assess their association in an independent replication dataset including 684 cases and 532 age-matched controls (Table 1). The association analyses in this dataset were performed adjusted for gender since the male to female ratio was significantly higher in the BD group when compared to the control group (49.3% versus 32.9%, respectively, *P*=1.81E-08).

Of the 47 genetic markers associated in the discovery dataset, 6 SNPs (3 in the *HLA-B* locus, and one in *DPH5*, *YY1P2*, and *FUT2*) were also significantly associated in the replication dataset (Supplementary Table 3 and Table 2). Association analyses of these 6 polymorphisms in the combined discovery and validation datasets (Table 2) strengthened the association observed at the *HLA-B* locus, a known risk locus for BD where rs6910516, rs9266490 and rs9266406 reached highly significant *P*-values (1.63E-44 ≤ $P_{combined}$ ≤ 2.97E-40). Outside the *HLA-B* region, rs7528842, located in a gene desert on chromosome 1p21.2 with the nearest gene being Diphthine synthase (*DPH5*) ($P_{combined}$=8.62E-04, OR_G[95%

CI]=1.35[1.13-1.62]) and rs632111, located in the 3'UTR of the α-(1,2) fucosyltransferase gene (*FUT2*) on chromosome 19p13.3 ($P_{combined}$=2.09E-05, OR_G[95%CI]=1.32[1.16-1.49]) were also associated in both datasets and in the combined analysis. On the other hand, rs10803575 which is located in a gene desert on chromosome 2q22.1, was associated in both the discovery and replication sample in opposite directions and therefore is not associated in the combined dataset ($P_{combined}$= 6.24E-01).

To follow-up on our novel findings (SNPs outside of the *HLA-B* locus), we checked the association of rs7528842 and rs632111 in a Turkish BD GWAS of effectively genotyped SNPs [Remmers *et al*., 2010] or of imputed data [Kirino *et al*., 2013]. In this group of 1215 BD cases and 1278 controls, rs632111 was associated with BD (*P*=1.11E-03 in the imputed genotypes dataset) but rs7528842 was not (*P*=8.68E-01 in the imputed genotypes dataset) (E. F. Remmers, personal communication), and therefore we decided to fine-map only the association of *FUT2*.

Table 2. BD association results for the SNPs associated in both the discovery and replication datasets. *P*-values from logistic regression using the log-additive model are shown unadjusted in the discovery sample and adjusted for gender in the replication and combined datasets. At each marker, the odds ratios (OR) and 95% confidence intervals (CI) refer to the risk allele in the discovery dataset and are shown only for significant *P*-values (*P*≤0.05).

dbSNP ID	Chr.	Gene	Dataset	Risk allele	Frequency		*P*-value	OR[95% CI]
					Cases	Controls		
rs7528842	1	-	Discovery	G	0.179	0.114	3.01E-03	1.67[1.18-2.35]
			Replication		0.204	0.173	4.33E-02	1.24[1.01-1.54]
			Combined		0.197	0.152	8.62E-04	1.35[1.13-1.62]
rs10803575	2	-	Discovery	G	0.470	0.358	1.91E-04	1.56[1.23-1.97]
			Replication		0.410	0.459	1.92E-02	0.82[0.70-0.97]
			Combined		0.428	0.423	6.24E-01	-
rs9266406	6	*HLA-B*	Discovery	A	0.507	0.258	3.45E-19	3.30[2.49-4.39]
			Replication		0.511	0.295	1.19E-25	2.56[2.12-3.08]
			Combined		0.510	0.282	5.37E-44	2.80[2.40-3.27]
rs9266490	6	*HLA-B*	Discovery	G	0.512	0.258	7.83E-19	3.15[2.40-4.14]
			Replication		0.520	0.291	2.28E-26	2.52[2.11-3.02]
			Combined		0.518	0.279	1.63E-44	2.73[2.35-3.17]
rs6910516	6	*HLA-B*	Discovery	G	0.491	0.258	1.87E-15	2.85[2.17-3.75]
			Replication		0.513	0.280	3.46E-25	2.51[2.09-3.02]
			Combined		0.507	0.272	2.97E-40	2.64[2.27-3.07]
rs632111	19	*FUT2*	Discovery	G	0.537	0.432	8.80E-04	1.46[1.17-1.82]
			Replication		0.556	0.492	8.00E-03	1.23[1.05-1.45]
			Combined		0.551	0.471	2.09E-05	1.32[1.16-1.49]

Chr.: Chromosome.

6.4.3 Fine-mapping of *FUT2*

To fine-map the association of *FUT2* with BD, 4 additional haplotype tagging SNPs (rs16982241, rs281377, rs601338, and rs602662) and 3 coding variants (rs492602, rs681343, and rs485186) (Supplementary Figure 1A) were genotyped in the full (discovery plus replication) dataset (976 BD cases and 826 controls). The association results in the combined dataset (Figure 2 and Table 3) reveal that, out of these 6 newly and successfully tested variants (rs485186 failed genotyping quality controls), the 5 exonic variants are strongly associated with BD (2.97E-06≤$P_{combined}$≤1.34E-04 in the log-additive model). All tested exonic variants in *FUT2* were associated with BD, including the A allele of rs601338 (*P*=1.18E-05, OR_A[95% CI]=1.34[1.17-1.52]) which encodes for a stop codon (W143X), and rs602662 (P=2.97E-06, OR_A[95% CI]=1.35[1.19-1.54]) which encodes for the G258S amino acid change. rs681343, a synonymous variant in high LD with rs601338 (r^2=0.95 in the Iranian control group, Supplementary Figure 1B), was also among the most associated variants (*P*=7.27E-06, OR_T[95% CI]=1.36[1.19-1.56]). A recessive model analysis for the non-synonymous variants shows an increased risk conferred by the homozygous genotype for both the non-functional W143X variant (*P*=2.24E-05, OR_{AA}[95%CI]=1.60[1.28-1.99]) and the missense G258S alteration (*P*=1.98E-06, OR_{AA}[95%CI]=1.67[1.36-2.08]).

Table 3. Association with BD of SNPs in *FUT2*. *P*-values from logistic regression using the log-additive model are shown adjusted for gender in the combined Iranian dataset. ORs and 95% CIs are shown only for significant *P*-values (*P*≤0.05).

dbSNP ID	Position (base pair)	Variant role (aminoacid change)	Allele	Allele frequency		Log-additive model	
				Cases	Controls	*P*-value	OR[95% CI]
rs16982241	49202859	Intronic	G	0.919	0.903	1.05E-01	-
rs492602	49206417	Coding-synonymous	C	0.534	0.459	4.86E-05	1.31[1.15-1.49]
rs681343	49206462	Coding-synonymous	T	0.519	0.435	7.27E-06	1.36[1.19-1.56]
rs281377	49206603	Coding-synonymous	C	0.603	0.533	1.34E-04	1.30[1.14-1.47]
rs601338	49206674	Non-sense (W143X)	A	0.535	0.454	1.18E-05	1.34[1.17-1.52]
rs602662	49206985	Missense (G258S)	A	0.550	0.465	2.97E-06	1.35[1.19-1.54]
rs632111	49208978	3' UTR	G	0.551	0.471	2.09E-05	1.32[1.16-1.49]

To assess if the *FUT2* association signals are independent, we performed a pairwise conditional analysis between all the *FUT2* associated SNPs with the log-additive model. After conditional analysis on either one of the four most associated SNPs (rs681343, rs601338, rs602662 or rs632111), all the associations become non-significant (Supplementary Table 4).

Since these SNPs are also in strong LD (r^2>0.84, Supplementary Figure 1B), we cannot pinpoint the actual risk variant using this approach.

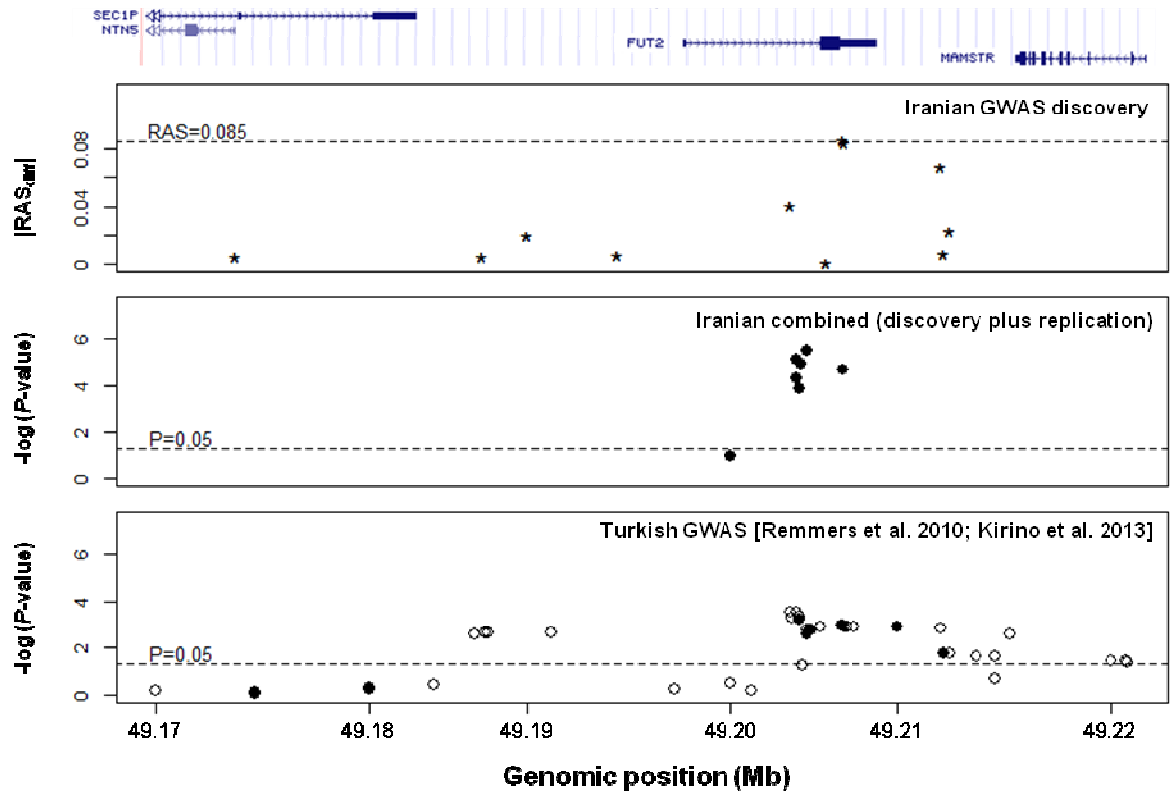


Figure 2. |RAS$_{diff}$| and association results in the *FUT2* genomic region. The top diagram depicts *FUT2* and its flanking genes (*SEC1P*, *NTN5*, *MAMSTR*). The first plot shows the |RAS$_{diff}$| for the 11 SNPs allelotyped in the *FUT2* genomic region with the Affymetrix platform in the discovery sample (292 BD cases and 294 controls). The second plot displays the association results (negative logarithm of the *P*-value) for the SNPs effectively genotyped in *FUT2* in the combined Iranian dataset (976 BD cases and 826 controls). The third plot depicts the association results of SNPs in the *FUT2* genomic region which were effectively genotyped (discs) or imputed (circles) in a Turkish dataset of 1215 BD cases and 1278 controls (E. F. Remmers, personal communication).

6.4.4 Meta-analyses in *FUT2*

To determine the risk conferred by *FUT2* variants in our full Iranian dataset and in a Turkish dataset [Remmers *et al.*, 2010; Kirino *et al.*, 2013], we performed a meta-analysis. It is important to notice that in the Turkish dataset some of these SNPs (rs281377 and rs602662) were effectively genotyped [Remmers *et al.*, 2010] while others (rs492602, rs681343, rs601338,

and rs632111) were imputed [Kirino *et al*., 2013] (Figure 2), but the imputation was of very high quality except for rs601338 which was then excluded from the meta-analysis (E. F. Remmers, personal communication).

The meta-analysis (Table 4) strengthened the associations of all *FUT2* variants tested ($4.78E\text{-}09 \leq P_{meta} \leq 1.66E\text{-}07$), reinforcing the role of *FUT2* in BD susceptibility. The nonsense rs601338 variant (W143X), which defines the ABO secretor phenotype in Caucasians, could not be meta-analysed but is in very high LD ($r^2 \geq 0.95$ in the Iranian and CEU, Supplementary Figures 1A and 1B) with the synonymous variants rs492602 and rs681343, suggesting that its association with BD would also be emphasized by the meta-analysis.

Table 4. Meta-analysis of five SNPs in *FUT2*. *P*-values from logistic regression using the log-additive model are shown adjusted for gender in the combined (discovery and replication) Iranian dataset. In the Turkish dataset, unadjusted allelic *P*-values from genotyped or imputed SNPs are indicated. The results of fixed effects (Mantel-Haenszel) meta-analyses are shown. ORs and 95% CIs are specified for all significant *P*-values ($P \leq 0.05$).

dbSNP ID	Risk allele	Dataset	Frequency		*P*-value	OR [95% CI]
			Cases	Controls		
rs492602	C	Iranian combined	0.534	0.459	4.86E-05	1.31[1.15-1.49]
		Turkish (imputation)	0.564	0.513	2.85E-04	1.23[1.10-1.37]
		Meta-analysis			3.05E-08	1.28[1.17-1.39]
rs681343	T	Iranian combined	0.519	0.435	7.27E-06	1.36[1.19-1.56]
		Turkish (imputation)	0.564	0.513	2.85E-04	1.23[1.10-1.37]
		Meta-analysis			4.78E-09	1.30[1.19-1.41]
rs281377	C	Iranian combined	0.603	0.533	1.34E-04	1.30[1.14-1.47]
		Turkish	0.613	0.566	6.25E-04	1.22[1.09-1.37]
		Meta-analysis			1.66E-07	1.26[1.16-1.38]
rs602662	A	Iranian combined	0.550	0.465	2.97E-06	1.35[1.19-1.54]
		Turkish	0.572	0.529	2.52E-03	1.19[1.06-1.33]
		Meta-analysis			3.96E-08	1.28[1.17-1.40]
rs632111	G	Iranian combined	0.551	0.471	2.09E-05	1.32[1.16-1.49]
		Turkish (imputation)	0.569	0.523	1.10E-03	1.20[1.45-1.07]
		Meta-analysis			1.12E-07	1.26[1.16-1.38]

6.5 DISCUSSION

In this study, we performed the first GWAS for Behçet's disease in an Iranian dataset, using a DNA pooling approach. We followed a three-stage design approach, in which the top hits from the pooled GWAS were confirmed by individual genotyping and the SNPs technically validated were further tested for association in an independent replication sample. Furthermore, meta-analysis with the results of another GWAS in a Turkish dataset strongly support the role of *FUT2* in BD susceptibility.

Our pooling strategy was technically sound since we identified the well-established association of the *HLA-B* locus with BD and validated by individual genotyping the association of 47 out of 51 tested SNPs (at the conventional *P*-value of 0.05). There is no general consensus on which analysis method (e.g. RAS difference, F ratio, silhouette score, cluster) is best to prioritize SNPs for individual genotyping in pooled GWAS. We choose the $|RAS_{diff}|$ to rank the top SNPs as it is a sensitive method to detect differences between cases and controls if the technical variation between pools is low. Conversely, as this method does not take into account RAS variation among replicates, it can lead to a high number of false positive and false negative results if the technical variation among pools is high. In our pooling design, to reduce the error attributed to pool construction (biological error) and to array differences (technical error), we constructed two biological replicates and performed four technical replicates, which resulted in low variability between arrays. We have also excluded SNPs with MAF<0.05 in order to reduce the number of false positives. It is therefore unlikely that high variance between replicates be the cause of the lack of validation for some SNPs.

In our study, the correlation between the RAS values and the true allele frequencies (r=0.853) was lower than what has been reported previously by others (r=0.943-0.969) [Abraham *et al*. 2008; Bossé *et al*. 2009; Bostrom *et al*. 2010]. This may derive from the fact that, in contrast to earlier versions of Affymetrix genotyping arrays, in the Genome-Wide SNP Array 6.0 used here, the mismatch probes have been discarded in favor of greater perfect match probe density. Thus, mismatch probe intensities cannot be subtracted before calculating RAS values and, as a consequence, RAS is no longer such a good estimate of the real allelic frequencies in the pool [Schosser *et al*., 2010]. Despite this, we unequivocally identified the *HLA-B* locus as the major genetic risk factor for Behçet's disease and identified an association of the *FUT2* gene that was replicated in independent Iranian and Turkish datasets.

FUT2 encodes an α-(1,2) fucosyltransferase that regulates the secretion of the H antigen (precursor of the human ABO blood group antigens) in body fluids and intestinal mucosa [Ferrer-Admetlla *et al*., 2009]. About 80% of individuals have the secretor phenotype which is determined by the presence of at least one functional *FUT2* allele, and the non-secretor occurs in the presence of two non-functional *FUT2* alleles [Kelly *et al*., 1995]. In subjects of European origin, rs601338 is the non-sense polymorphism that most commonly determines the secretor status (its functional G allele encodes a tryptophan at position 143 of the protein, while the non-functional A allele encodes a stop codon at the same position). The non-secretor phenotype was strongly associated with BD in the combined Iranian dataset but unfortunately it was not effectively genotyped or accurately imputed in the Turkish. However, a *FUT2* missense SNP (rs602662) in very strong LD with rs601338 (r^2=0.92) which was among the most associated markers in Iranian and Turkish individually and combined, has also been associated with decreased or absent FUT2 enzyme activity [Serpa *et al*., 2004] and may also be responsible for some instances of non-secretor status. Since the investigated *FUT2* coding variants are located in a tight haplotype block, functional studies are warranted to determine exactly which functional variant explains the increased risk for Behçet's disease.

The highly variable incidence of BD worldwide, migrant studies [Zouboulis *et al*. 1997, Sakane *et al*. 1999], and the beneficial effect of antibacterial treatments on mucocutaneous and synovial disease symptoms [Mumcu *et al*. 2007] strongly support the existence of environmental factors. The search for bacterial or viral triggers for BD has not yet been conclusive, and one possibility is that it may depend on a gene-environment interaction. Indeed, the nonsecretor *FUT2* phenotype has been associated with reduced susceptibility to infections by the Norwalk virus [Lindesmith *et al*. 2003], Campylobacter jejuni [Ruiz-Palacios *et al*. 2003], and Helicobacter pylori [Ikehara *et al*. 2001] and with an increased risk for inflammatory and/or auto-immune disorders such as rheumatic fever [Haverkorn *et al*., 1969], Crohn's disease [McGovern *et al*., 2010] and type I diabetes (T1D) [Smyth *et al*., 2011]. Based on the observation that FUT2 nonsecretors show alterations in the mucosal glycosylation and gut microbiome [Magalhães *et al*., 2011; Wacklin *et al*., 2011], Yang *et al*. [2011] proposed that these individuals may have an increase risk for T1D due to lower antigenic stimulation in early life and consequent increased propensity for auto-immunity. A similar mechanism may be at work in BD pathogenesis. Namely, future investigation may be the impact of the FUT2 secretor status on the oral and gastrointestinal tract flora, on the development of the immune system and inflammation, in particular at the level of blood vessels.

Our study highlighted *FUT2* as a novel genetic risk factor for Behçet's disease and possibly as a missing link in the gene-environment interaction in BD. More studies will be necessary to better understand how the secretor phenotype encoded by *FUT2* modulates susceptibility to BD risk, if by changing the gut microbiome and influencing resistance to a bacterial trigger or by any other mechanism.

6.6 REFERENCES

Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics. 2008;1:44.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

Bossé Y, Bacot F, Montpetit A, Rung J, Qu HQ, Engert JC, Polychronakos C, Hudson TJ, Froguel P, Sladek R, Desrosiers M. Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. Hum Genet. 2009;125(3):305-18.

Bostrom MA, Lu L, Chou J, Hicks PJ, Xu J, Langefeld CD, Bowden DW, Freedman BI. Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA. Hum Genet. 2010;128(2):195-204.

Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, Puffenberger EG, Pearson JM, Stephan DA. Identification of disease causing loci using an array-based genotyping approach on pooled DNA. BMC Genomics. 2005;6:138.

Davis OS, Plomin R, Schalkwyk LC. The SNPMaP package for R: A framework for genome-wide association using DNA pooling on microarrays. Bioinformatics. 2009;25(2):281-3.

Fei Y, Webb R, Cobb BL, Direskeneli H, Saruhan-Direskeneli G, Sawalha AH. Identification of novel genetic susceptibility loci for Behçet's disease using a genome-wide association study. Arthritis Res Ther. 2009;11(3):R66.

Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, Calafell F, Bertranpetit J, Casals F. A natural history of FUT2 polymorphism in humans. Mol Biol Evol. 2009;26(9):1993-2003.

Gaj P, Maryan N, Hennig EE, Ledwon JK, Paziewska A, Majewska A, Karczmarski J, Nesteruk M, Wolski J, Antoniewicz AA, Przytulski K, Rutkowski A, Teumer A, Homuth G, Starzyńska T, Regula J, Ostrowski J. Pooled sample-based GWAS: a cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population. PLoS One. 2012;7(4):e35307.

González JR, Armengol L, Solé X, Guino E, Mercader JM, Estivill X, Moreno V. SNPassoc: an R package to perform whole genome association studies. Bioinformatics 2007;23:644-5.

Haverkorn MJ, Goslings WR. Streptococci, ABO blood groups, and secretor status. Am J Hum Genet. 1969;21(4):360-75.

Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, Li F, Zhou Q, Ohno S, Chen R, Kijlstra A, Rosenbaum JT, Yang P. Identification of a susceptibility locus in STAT4 for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum. 2012;64(12):4104-13.

Ikehara Y, Nishihara S, Yasutomi H, Kitamura T, Matsuo K, Shimizu N, Inada K, Kodera Y, Yamamura Y, Narimatsu H, Hamajima N, Tatematsu M. Polymorphisms of two

fucosyltransferase genes (Lewis and Secretor genes) involving type I Lewis antigens are associated with the presence of anti-Helicobacter pylori IgG antibody. Cancer Epidemiol Biomarkers Prev. 2001;10(9):971-7.

International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD). Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 2006;24(suppl 42): S14-S5.

Kaya Tİ. Genetics of Behçet's Disease. Patholog Res Int. 2012;2012:912589.

Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem. 1995;270(9):4640-9.

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sacli FS, Erer B, Inoko H, Emrence Z, Cakar A, Abaci N, Ustek D, Satorius C, Ueda A, Takeno M, Kim Y, Wood GM, Ombrello MJ, Meguro A, Gül A, Remmers EF, Kastner DL. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013;45(2):202-7.

Lee YH, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of a genome-wide association study on psoriasis and Behcet's disease. Mol Biol Rep. 2012;39(5):5953-9.

Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindblad L, Stewart P, LePendu J, Baric R. 2003. Human susceptibility and resistance to Norwalk virus infection. Nat Med. 9:548–553.

Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res. 2008;36(6):e35.

Magalhães A, Gomes J, Ismail MN, Haslam SM, Mendes N, Osório H, David L, Le Pendu J, Haas R, Dell A, Borén T, Reis CA. Fut2-null mice display an altered glycosylation profile and impaired BabA-mediated Helicobacter pylori adhesion to gastric mucosa. Glycobiology. 2009;19(12):1525-36.

McGovern DP, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, Ippoliti A, Vasiliauskas E, Berel D, Derkowski C, Dutridge D, Fleshner P, Shih DQ, Melmed G, Mengesha E, King L, Pressman S, Haritunians T, Guo X, Targan SR, Rotter JI; International IBD Genetics Consortium. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum Mol Genet. 2010;19(17):3468-76.

Melquist S, Craig DW, Huentelman MJ, Crook R, Pearson JV, Baker M, Zismann VL, Gass J, Adamson J, Szelinger S, Corneveaux J, Cannon A, Coon KD, Lincoln S, Adler C, Tuite P, Calne DB, Bigio EH, Uitti RJ, Wszolek ZK, Golbe LI, Caselli RJ, Graff-Radford N, Litvan I, Farrer MJ, Dickson DW, Hutton M, Stephan DA (2007). Identification of a novel risk locus for progressive supranuclear

palsy by a pooled genomewide scan of 500,288 single-nucleotide polymorphisms. Am J Hum Genet. 2007;80(4):769-78

Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, Ito N, Kera J, Okada E, Yatsu K, Song YW, Lee EB, Kitaichi N, Namba K, Horie Y, Takeno M, Sugita S, Mochizuki M, Bahram S, Ishigatsubo Y, Inoko H. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat Genet. 2010 Aug;42(8):703-6.

Mumcu G, Inanc N, Yavuz S, Direskeneli H. The role of infectious agents in the pathogenesis, clinical manifestations and treatment strategies in Behçet's disease. Clin Exp Rheumatol. 2007;25(4 Suppl 45): S27-33.

Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, Heun R, Jessen F, Kolsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. Am J Hum Genet. 2007;80(1):126-39.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007;81(3):559-75.

Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, Le JM, Yang B, Korman BD, Cakiris A, Aglar O, Emrence Z, Azakli H, Ustek D, Tugal-Tutkun I, Akman-Demir G, Chen W, Amos CI, Dizon MB, Kose AA, Azizlerli G, Erer B, Brand OJ, Kaklamani VG, Kaklamanis P, Ben-Chetrit E, Stanford M, Fortune F, Ghabra M, Ollier WE, Cho YH, Bang D, O'Shea J, Wallace GR, Gadina M, Kastner DL, Gül A. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet. 2010;42(8):698-702.

Ricci G, Astolfi A, Remondini D, Cipriani F, Formica S, Dondi A, Pession A. Pooled genome-wide analysis to identify novel risk loci for pediatric allergic asthma. PLoS One. 2011;6(2):e16912.

Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. J Biol Chem. 2003;278(16):14112-20.

Sakane T, Takeno M, Suzuki N, Inaba G. Behçet's disease. N Engl J Med. 1999;341(17): 1284-91.

Schosser A, Pirlo K, Gaysina D, Cohen-Woods S, Schalkwyk LC, Elkin A, Korszun A, Gunasinghe C, Gray J, Jones L, Meaburn E, Farmer AE, Craig IW, McGuffin P. Utility of the pooling approach as applied to whole genome association scans with high-density Affymetrix microarrays. BMC Res Notes. 2010;3:274.

Serpa J, Mendes N, Reis CA, Santos Silva LF, Almeida R, Le Pendu J, David L. Two new FUT2 (fucosyltransferase 2 gene) missense polymorphisms, 739G-->A and 839T-->C, are partly

responsible for non-secretor status in a Caucasian population from Northern Portugal. Biochem J. 2004;383(Pt. 3):469-74.

Smyth DJ, Cooper JD, Howson JM, Clarke P, Downes K, Mistry T, Stevens H, Walker NM, Todd JA. FUT2 nonsecretor status links type 1 diabetes susceptibility and resistance to infection. Diabetes. 2011;60(11):3081-4.

Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR. Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. Genes Immun. 2007 Jan;8(1):57-68.

Wacklin P, Mäkivuokko H, Alakulppi N, Nikkilä J, Tenkanen H, Räbinä J, Partanen J, Aranko K, Mättö J. Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. PLoS One. 2011;6(5):e20113.

Yang P, Li HL, Wang CY.FUT2 nonfunctional variant: a "missing link" between genes and environment in type 1 diabetes? Diabetes. 2011;60(11):2685-7.

Zouboulis CC, Kötter I, Djawari D, Kirch W, Kohl PK, Ochsendorf FR, Keitel W, Stadler R, Wollina U, Proksch E, Söhnchen R, Weber H, Gollnick HP, Hölzle E, Fritz K, Licht T, Orfanos CE. Epidemiological features of Adamantiades-Behçet's disease in Germany and in Europe. Yonsei Med J. 1997;38(6): 411-22.

6.7 SUPPLEMENTARY MATERIAL

Supplementary Table 1. Primer sequences used to effectively genotype 57 SNPs in the Iranian dataset.

dbSNP ID	PCR primer 1	PCR primer 2	Extension primer
rs16835990	ACGTTGGATGCCTCCCCTCATTCCATTTTG	ACGTTGGATGAGGGAAGTAGAAGTAGGGAG	CGTAGGGAGAATCTACCTGAATAATT
rs3737741	ACGTTGGATGAGTGAGTGTGACCTTCACAG	ACGTTGGATGGAGGTGGGTTCTATGATTAG	GGTTTCTATGATTAGCCCCAC
rs2994552	ACGTTGGATGAGTCTTTCCCTGCAGAGAAG	ACGTTGGATGGGTGGCCGCAGGAATTATTG	GCAGGAATTATTGTATGATCTAC
rs7528842	ACGTTGGATGTATGAAAGGGTGGAGGGAAG	ACGTTGGATGCATGTTTCTTTAGTCTCTTG	TGTTTCTTTAGTCTCTTGTAATCTA
rs569659	ACGTTGGATGTTACCCTCTGCCCAAGGAAG	ACGTTGGATGACTATGAAGCTCCCCATCAG	TCCCTCCTCACTACTG
rs11584700	ACGTTGGATGTGGGTTTCAGTTGTCTAAGG	ACGTTGGATGTAAGTGACAGAAAGCACATC	CCCCCCCCGTCCAGCTTAACAA
rs9753377	ACGTTGGATGAACAGACCACCTTTGCTGAG	ACGTTGGATGATTGTCAGCCCAGCAACTTC	CCCCCTCTTCTAAAGCAAGACTGTC
rs10928683	ACGTTGGATGGGTAGCTAGATTATGTCTGG	ACGTTGGATGGGCTACATATGAGAGAAACC	GATGAGAGAAACCTATAAGAGA
rs10803575	ACGTTGGATGCATCCTCTGTAAAAACAAAC	ACGTTGGATGTAGCCTGGGACATTCTAGAG	GGTCGCCTGGGACATTCTAGAGTTAGAC
rs10803657	ACGTTGGATGTCCATGTCTCCTTCCATCTC	ACGTTGGATGATACTGGTGCTGGTTCTTCC	TGGTTCTTCCATAACTAATAACT
rs11130135	ACGTTGGATGGAGGTACCTCCATTTTACAG	ACGTTGGATGACTGCCTCCTAGCTGCATGT	GCTGCATGTCCGTAA
rs9311918	ACGTTGGATGCCCATTCTGTTACTAAGAAA	ACGTTGGATGAAAAACCATGGATCTGTGT	TGGATCTGTGTAGATGC
rs1477674	ACGTTGGATGCGGATGGGAGATCAGCATAG	ACGTTGGATGTGCTGAGAGTAACATATCTG	GAGAGTAACATATCTGTAACTT
rs16838396	ACGTTGGATGGGCATTGTGTTTAGTTTGTG	ACGTTGGATGTAGAGTCAATGCTCATAGGG	TTTTATCTTTGGCCCTACAT
rs9847124	ACGTTGGATGGAGATGCAGACTAAGCTTTG	ACGTTGGATGGTTACAATGAATGATCTCTG	GAGTGTGAGTTGGTTTCTAAGATCATAA
rs207340	ACGTTGGATGCATGTTTTAAAGCTAGTTC	ACGTTGGATGCTTTAATTGACCTAAGTAG	AAAAATTTTTGAATTCCCTAGAT
rs10000286	ACGTTGGATGAGGGCAAACCAAATGCAAAC	ACGTTGGATGTTACCCCAATCTCTCTATCC	GGAACCCAATCTCTCTATCCTACTAAAA
rs11728991	ACGTTGGATGGCTTTAGGCAACCAGTAATC	ACGTTGGATGCGGTCACAAAGAAGACATAT	GAAGACATATTGTCTGATTCAATTTATA
rs4336313	ACGTTGGATGACTCCCCTGATCTGTGATTC	ACGTTGGATGAACAACACCCCACATGGAAG	CCTGCTTTGAGCAACA
rs2931429	ACGTTGGATGAGCAATTCCTAGGCTAACGC	ACGTTGGATGCTGTCCCTCACTTCATTCTG	ATTCTGACATACTTTTTCATGG
rs10474500	ACGTTGGATGAAACCACCTCCCAGCAATAG	ACGTTGGATGGGGTGCAGAAGAATATTGAG	GGGGAGAAGAATATTGAGTGCCTTAAC
rs7731597	ACGTTGGATGGAGACAGGGAATAGCAAGTG	ACGTTGGATGAAGAAGCACATGTCTGGAGC	GCAGAAGGATTAATCTTACCAGC
rs9266406	ACGTTGGATGTACTCAGTCACAAACCACCC	ACGTTGGATGTCAACTGGAACTGGCAGAAC	TCTGTGTTAGAAATCAGGA
rs9266490	ACGTTGGATGATTGACTGGAGGGCTTCAAG	ACGTTGGATGTCTGCTCCTCAAATGGGCAA	GAGGGATCAGTAAACGCAAACAT
rs6910516	ACGTTGGATGGGAATGTATTATTATACAAG	ACGTTGGATGCATCTTCTTATCTGAAATG	ATGTATTATTATACAAGTAGTTTTCC
rs6937876	ACGTTGGATGAAATAGGGTCGCACAGACAG	ACGTTGGATGTTCCTTAACTCCAGGTGGAC	GCATAGACACTCATGGATGTA
rs2285430	ACGTTGGATGTCCTCCTTTGTTTTCATGGC	ACGTTGGATGATTCAAAGCTCCTGTGACCC	GTTTTTTTCATGGCAAAACATACTA
rs6472266	ACGTTGGATGCTAGTATTGTAAGGGAATT	ACGTTGGATGCTGTTTTATGATGGCCTTAC	TGGCCTTACAGGAGA

rs7832811	ACGTTGGATGCCTACTCAATGGAACTCTTC	ACGTTGGATGAGGATGTTTTGAAATACAC	AAATTTACTTCATAATTCCTGTTAA
rs888254	ACGTTGGATGAACTCCTCTTGCAGTTTTAC	ACGTTGGATGGCTAGACAACTCTCTGAATG	GAGCCTGTCCAGATAATGTTCTAGAAT
rs2149879	ACGTTGGATGAATTTGTAACCCTCGGGCTC	ACGTTGGATGCCCTATCTCGTTTTACAGAC	GCCATATCTCGTTTTACAGACTTTATGA
rs303573	ACGTTGGATGATCAGATTGTCTGAGCTCCC	ACGTTGGATGAACCAGGAGCCTGACTTTAG	CCCCCAGACTTACCTGACATA
rs4749421	ACGTTGGATGGGGTAAATATTTGTCTAAGGC	ACGTTGGATGAGTAGCCAGAATTAAGAGGG	ACTGATTCCAGTGGAAA
rs4749423	ACGTTGGATGGGAGTTGTGGATGAGGATAG	ACGTTGGATGGAGTGGTTCTGAGCTGTTTG	TCTTAGCTGTTTGATTTAGTCAAACT
rs10824019	ACGTTGGATGCCCGCTTACATGAATGACTG	ACGTTGGATGAGACACCCTGATGTTTCTCC	TGATGTTTCTCCAATCTTCTAT
rs878824	ACGTTGGATGGCTGCTTGTTAAGAGCTTAG	ACGTTGGATGGGGCTTGTCATCTCATGTGC	TCTCATGTGCATTCCAG
rs10786172	ACGTTGGATGAGGCCATTTTGAACAGCAAG	ACGTTGGATGCCTACTCCAGCATTTGAACC	GAACCTGAAGTTTTGAGAA
rs7936044	ACGTTGGATGGTATCAATCCCTTCTCTGAC	ACGTTGGATGTGGGAGTCAATGGGTTAGTG	GGACTTGGGCTACAGCTC
rs497279	ACGTTGGATGCTTGCGTGTGTTCATACAGC	ACGTTGGATGGTGACTTGGTCACTATTTAG	TTCATACAGCTGTTCATTATAG
rs552221	ACGTTGGATGTGTACAGAGAGGCAAGTAGG	ACGTTGGATGGATGGTTGCAACCTACAGTC	CAACCTACAGTCAGGTTAC
rs17494112	ACGTTGGATGTAGGCCAGCACAGAACAAAG	ACGTTGGATGGAGCACTTCATAGGTATGCG	GGTGTGCGAAGGATTTGCTGTTCA
rs3852594	ACGTTGGATGGGCCCATGGTAATGAATAGG	ACGTTGGATGGTGTAGTAGACTGTATTACTG	TGTTCAGTTATTCACTCCC
rs6538225	ACGTTGGATGTCACCACTGAATTGGCTTAC	ACGTTGGATGTTTGGTTGCAATCTCAACAG	TTGTTTGCAATCTCAACAGATACTAT
rs7206095	ACGTTGGATGGGACACATGGTGATTATTTG	ACGTTGGATGTCAGGGTGGTGATTGTTGAG	TGTGGTGATTGTTGAGTAAATCAAA
rs16982241	ACGTTGGATGCTGGGAAACCTGGGTTCTTC	ACGTTGGATGTGCTATTGTCACCCACAAGG	GGCAAAGAATGCTGTT
rs492602	ACGTTGGATGAACATCAAAGGCACTGGGAC	ACGTTGGATGCCATCTGGTTCCCCAGGCG	GGGCTGTGGACGATCAATGC
rs681343	ACGTTGGATGAACATCAAAGGCACTGGGAC	ACGTTGGATGCCATCTGGTTCCCCAGGCG	GGGCTGTGGACGATCAATGC
rs281377	ACGTTGGATGATCCCCTGGCAGAACTACC	ACGTTGGATGTAGCCGGTGAAGCGGACGTA	GCAGAACTACCACCTGAA
rs601338	ACGTTGGATGTACGTCCGCTTCACCGGCTA	ACGTTGGATGAACTCCTGGAGGATCTCCTG	GCTACCCCTGCTCCT
rs602662	ACGTTGGATGCTGGGAAACCTGGGTTCTTC	ACGTTGGATGTGCTATTGTCACCCACAAGG	GGCAAAGAATGCTGTT
rs632111	ACGTTGGATGGGAATACCCAACTCCATGAC	ACGTTGGATGCCTCTCAGAGGATACATTTG	CCTCACCATTACACAGTCATCA
rs16983549	ACGTTGGATGATCTTGCGTGAGTTGGGTTC	ACGTTGGATGAGGAGAGACCATCACAACAC	AACTCAGATGCTCATCGTC
rs10417820	ACGTTGGATGACGAAGACTTATGACACCAC	ACGTTGGATGGGACTGTTTCTGTACCCTTG	ACCCTTGATTCCATTCTAT
rs7256125	ACGTTGGATGCAGAACTTCCTGTTGCTGTC	ACGTTGGATGGCTGGCAACTGCTATATCAC	GGAGACTTAGAAAAGTTAACGATT
rs685119	ACGTTGGATGAATTCCATCCTGTGGAGGAG	ACGTTGGATGGGGTTATGTCATAAGCCTTG	CTTCTGTCATAAGCCTTGACAAGA
rs7290471	ACGTTGGATGTGAACTCCTGACCTCAGGTG	ACGTTGGATGTGGCTTATTTGCTTGTGTGC	CATCTGTGTACTATTAATCACTGTC
rs2854131	ACGTTGGATGACAAACCCTGTTCTTGGGTG	ACGTTGGATGTACTTCACAAAGCGCCTTCC	GTAAATGATATCATCTCAACTTAG

Supplementary Table 2. **SNPs with $|RAS_{diff}| \geq 8.5\%$ in the Behçet's disease GWAS in pools.** The probe set ID, dbSNP ID, chromosome (Chr.), base pair position (bp), nearest gene, minor allele frequencies (MAF) in the CEU HapMap population, $|RAS_{diff}|$ and the alleles existing in the HapMap individuals, are indicated for all SNPs.

Probe set ID	dbSNP ID	Chr.	Position (pb)	Gene	MAF	$\|RAS_{diff}\|$	Alleles
SNP_A-1893234	rs6910516	6	31343827	*HLA-B*	0.212	0.173	A:G
SNP_A-8284151	rs10504160	8	55013096	*LYPLA1*	0.004	0.152	C:T
SNP_A-4273175	rs9266406	6	31336418	*HLA-B*	0.212	0.148	G:A
SNP_A-4207170	rs2919427	16	30669091	*PRR14*	0	0.142	A:G
SNP_A-8714283	rs9266490	6	31340158	*HLA-B*	0.218	0.141	A:G
SNP_A-8639063	rs9380217	6	31051553	*HCG22*	0.062	0.140	C:T
SNP_A-4253934	rs4959053	6	31099577	*PSORS1C1*	0.088	0.140	G:A
SNP_A-4219514	rs5022119	6	31343862	*HLA-B*	0.35	0.136	T:A
SNP_A-2025367	rs6933050	6	31343632	*HLA-B*	0.198	0.133	T:C
SNP_A-8574715	rs2854131	MT	3197	*MT-RNR2*	NA	0.127	T:C
SNP_A-4239246	rs4947296	6	31058178	*C6orf15*	0.062	0.122	T:C
SNP_A-2167622	rs2523638	6	31344273	*HLA-B*	0.429	0.121	T:C
SNP_A-4302686	rs9266409	6	31336568	*HLA-B*	0.212	0.119	T:C
SNP_A-1819202	rs7256125	19	52742395	*PPP2R1A*	0.128	0.117	A:G
SNP_A-4260888	rs2844558	6	31340433	*HLA-B*	NA	0.117	C:T
SNP_A-4268470	rs6538225	12	77916775	*NAV3*	0.204	0.117	G:T
SNP_A-1837957	rs5025315	6	31343604	*HLA-B*	0.344	0.114	A:G
SNP_A-8611784	rs2853943	6	31247871	*HLA-B*	0.221	0.110	T:C
SNP_A-8698987	rs5006725	6	31337872	*HLA-B*	0.333	0.109	A:G
SNP_A-4235979	rs10803575	2	139800877	*LOC647012*	0.487	0.108	T:G
SNP_A-8563168	rs5006724	6	31337850	*HLA-B*	0.345	0.108	T:C
SNP_A-2178269	rs3094584	6	31383848	*MICA*	0.142	0.107	G:A
SNP_A-2073652	rs2523534	6	31336349	*HLA-B*	0.367	0.106	A:G
SNP_A-8449009	rs2253907	6	31336870	*HLA-B*	0.436	0.106	C:T
SNP_A-8358852	rs16851566	1	163232649	*RGS5*	0.035	0.105	T:C
SNP_A-1880604	rs2073716	6	31122997	*CCHCR1*	0.031	0.105	C:G
SNP_A-2191746	rs10417820	19	52726346	*PPP2R1A*	0.121	0.105	A:G
SNP_A-8685280	rs6937876	6	106580629	*PRDM1*	0.434	0.104	G:A
SNP_A-2242006	rs4749421	10	29711520	*LOC387647*	0.248	0.098	C:T
SNP_A-8649222	rs4954974	2	139670767	*NXPH2*	0.447	0.097	G:T
SNP_A-8463850	rs2931429	5	73228106	*RGNEF*	0.088	0.097	C:A
SNP_A-8427443	rs1355282	6	124375684	*NKAIN2*	0.407	0.097	A:T
SNP_A-8624394	rs10786172	10	96581094	*CYP2C19*	0.376	0.096	A:G
SNP_A-1851889	rs10474500	5	76620002	*PDE8B*	0.261	0.096	G:A
SNP_A-8711469	rs878824	10	82795855	*SH2D4B*	0.125	0.096	A:G
SNP_A-8668957	rs2285430	7	18674040	*HDAC9*	0.425	0.095	T:G
SNP_A-2141731	rs3737741	1	47000566	*KNCN*	0.478	0.095	T:C
SNP_A-8475213	rs11130135	3	4809618	*ITPR1*	0.102	0.094	A:G
SNP_A-8347124	rs2994552	1	56787046	*PPAP2B*	0.326	0.094	A:C
SNP_A-2097132	rs303573	9	106103152	*CYLC2*	0.088	0.092	T:C
SNP_A-2079423	rs2523467	6	31362930	*MICA*	0.274	0.092	C:T
SNP_A-8487230	rs685119	20	813751	*FAM110A*	0.478	0.092	C:G
SNP_A-8328360	rs207340	4	19044103	*LCORL*	0.363	0.092	G:A
SNP_A-8622125	rs4336313	5	68320768	*SLC30A5*	0.144	0.092	C:T
SNP_A-4286652	rs10803657	2	238214039	*COL6A3*	0.367	0.091	C:T
SNP_A-2050588	rs17494112	12	42355218	*YAF2*	0.173	0.091	C:G
SNP_A-2092038	rs10928683	2	139763141	*LOC647012*	0.4	0.091	C:G
SNP_A-2236975	rs10000286	4	32037931	*PCDH7*	0.199	0.090	G:A
SNP_A-8318456	rs16838396	3	131756012	*CPNE4*	0.058	0.090	G:A
SNP_A-4300608	rs2596464	6	31412961	*HCP5*	0.487	0.089	A:G

SNP_A-8617817	rs9311918	3	64833413	*MIR548A2*	0.314	0.089	A:G
SNP_A-8711315	rs888254	8	113275266	*CSMD3*	0.226	0.089	C:G
SNP_A-8462189	rs9753377	2	1300591	*SNTG2*	0.416	0.088	C:T
SNP_A-8380904	rs552221	11	131784072	*NTM*	0.226	0.088	C:T
SNP_A-8516685	rs2149879	9	92128645	*SEMA4D*	0.341	0.088	A:G
SNP_A-8633511	rs11584700	1	204576983	*LRRN2*	0.208	0.088	A:G
SNP_A-1860842	rs7832811	8	111519986	*SYBU*	0.288	0.088	A:G
SNP_A-4265230	rs7731597	5	135969451	*TRPC7*	0.232	0.088	G:T
SNP_A-4225863	rs16983549	19	52703259	*PPP2R1A*	0.185	0.087	T:C
SNP_A-8367050	rs16835990	1	34314519	*CSMD2*	0.15	0.087	C:T
SNP_A-2008426	rs7936044	11	21666526	*NELL1*	0.416	0.087	T:C
SNP_A-1843198	rs497279	11	89217708	*NOX4*	0.243	0.087	T:C
SNP_A-8417606	rs11728991	4	68084081	*CENPC1*	0.186	0.087	A:G
SNP_A-1860507	rs3852594	12	42528784	*GXYLT1*	0.15	0.087	C:G
SNP_A-4256594	rs7206095	16	13279676	*SHISA9*	0.084	0.087	G:A
SNP_A-2126646	rs12364790	11	28890240	*METTL15*	0.257	0.087	T:C
SNP_A-8421823	rs7290471	22	50814524	*SAPS2*	0.381	0.086	A:G
SNP_A-2121118	rs7528842	1	101583663	*DPH5*	0.239	0.086	T:G
SNP_A-1851162	rs7577271	2	225709926	*DOCK10*	0.111	0.086	A:G
SNP_A-1820893	rs9847124	3	152590762	*P2RY1*	0.102	0.086	T:G
SNP_A-1848654	rs4749423	10	29717839	*LOC387647*	0.248	0.086	A:G
SNP_A-8402545	rs569659	1	160790411	*LY9*	0.496	0.086	A:G
SNP_A-2147879	rs6472266	8	67227072	*RRS1*	0.31	0.086	A:T
SNP_A-8606607	rs632111	19	49208978	*FUT2*	0.429	0.085	A:G
SNP_A-8673598	rs3130942	6	31197293	*HCG27*	0.252	0.085	C:T
SNP_A-4259036	rs1592050	10	93541969	*TNKS2*	0.226	0.085	A:T
SNP_A-8347367	rs1477674	3	81238571	*GBE1*	0.19	0.085	G:A
SNP_A-1934230	rs10824019	10	53834086	*PRKG1*	0.274	0.085	T:C

Supplementary Table 3. Association results of the GWAS SNPs selected for technical validation. The dbSNP ID, chromosome (Chr.), nearest gene, risk allele, Hardy Weinberg equilibrium (HWE) *P*-value in the control group, frequency (freq.) in cases and controls in the discovery dataset and *P*-value in the discovery dataset are shown for all the SNPs. Odds ratios (OR) and 95% confidence intervals (CI) are only shown for significant *P*-values ($P \leq 0.05$). *P*-values from logistic regression using the log-additive model are shown unadjusted for the GWAS discovery dataset and adjusted for gender for the replication dataset. Only SNPs associated in the discovery dataset ($P \leq 0.05$) were tested for association in the replication sample.

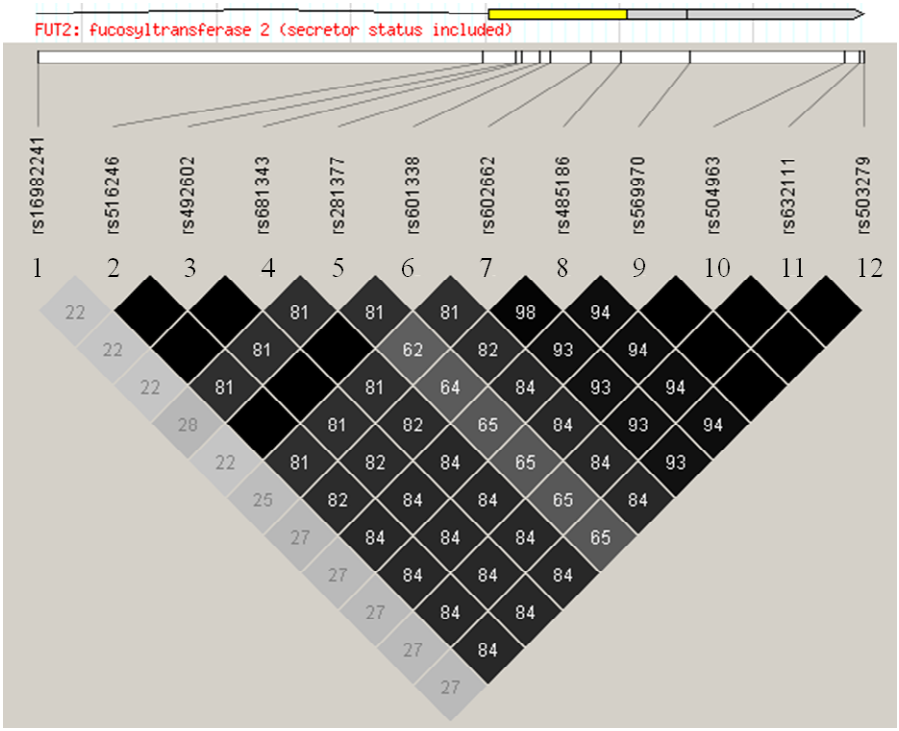
dbSNP ID	Chr.	Gene	Allele	Discovery					Replication				
				HWE *P* controls	Case freq	Control freq	P_{unadj}	OR [95% CI]	HWE *P* controls	Case freq	Control freq	P_{adj}	OR [95% CI]
rs16835990	1	*CSMD2*	C	0.935	0.828	0.767	9.84E-03	1.47[1.10-1.96]	0.245	0.191	0.170	1.88E-01	
rs3737741	1	*KNCN*	T	0.405	0.604	0.479	3.94E-05	1.61[1.28-2.04]	0.893	0.554	0.541	3.26E-01	
rs2994552	1	*PPAP2B*	A	0.152	0.783	0.716	1.06E-02	1.41[1.09-1.85]	0.086	0.276	0.270	6.35E-01	
rs7528842	1	*DPH5*	G	0.186	0.179	0.114	3.01E-03	1.67[1.18-2.35]	0.179	0.204	0.173	4.33E-02	1.24[1.01-1.54]
rs569659	1	*LY9*	A	1.000	0.458	0.327	8.63E-06	1.72[1.35-2.19]	0.864	0.392	0.365	2.54E-01	
rs11584700	1	*LRRN2*	A	0.848	0.803	0.710	3.27E-04	1.64[1.25-2.13]	0.453	0.253	0.247	5.98E-01	
rs9753377	2	*SNTG2*	C	0.728	0.507	0.482	3.99E-01	-	-	-	-	-	-
rs10928683	2	*LOC647012*	G	0.752	0.316	0.228	1.26E-03	1.52[1.18-1.97]	0.011	0.718	0.691	1.39E-01	
rs10803575	2	*LOC647012*	G	0.859	0.470	0.358	1.91E-04	1.56[1.23-1.97]	0.141	0.410	0.459	1.92E-02	0.82[0.7-0.97]
rs10803657	2	*COL6A3*	C	0.365	0.381	0.269	6.31E-05	1.67[1.29-2.15]	0.155	0.668	0.651	5.03E-01	
rs11130135	3	*ITPR1*	G	0.333	0.112	0.072	2.13E-02	1.60[1.07-2.42]	0.947	0.113	0.103	4.17E-01	
rs9311918	3	*MIR548A2*	G	0.795	0.349	0.255	7.43E-04	1.55[1.20-2.02]	0.074	0.300	0.293	5.68E-01	
rs1477674	3	*GBE1*	G	0.984	0.307	0.244	1.86E-02	1.36[1.05-1.77]	0.854	0.718	0.713	8.14E-01	
rs16838396	3	*CPNE4*	A	1.000	0.077	0.057	1.70E-01	-	-	-	-	-	-
rs9847124	3	*P2RY1*	T	0.603	0.833	0.755	1.68E-03	1.56[1.18-2.08]	0.366	0.778	0.776	8.08E-01	
rs207340	4	*LCORL*	G	0.191	0.356	0.320	2.29E-01	-	-	-	-	-	-
rs10000286	4	*PCDH7*	A	0.047	0.657	0.570	3.45E-03	1.41[1.12-1.79]	0.823	0.400	0.385	7.13E-01	
rs11728991	4	*CENPC1*	A	0.893	0.820	0.719	4.30E-05	1.82[1.35-2.44]	0.310	0.222	0.215	6.03E-01	
rs4336313	5	*SLC30A5*	T	0.915	0.757	0.663	4.77E-04	1.59[1.22-2.04]	0.741	0.291	0.261	1.30E-01	
rs2931429	5	*RGNEF*	C	1.000	0.089	0.037	5.73E-04	2.60[1.53-4.42]	0.364	0.066	0.054	3.15E-01	
rs10474500	5	*PDE8B*	G	0.427	0.239	0.194	6.55E-02	-	-	-	-	-	-
rs7731597	5	*TRPC7*	T	0.041	0.860	0.796	7.21E-03	1.49[1.11-2.00]	1.000	0.820	0.811	6.38E-01	
rs9266406	6	*HLA-B*	A	0.577	0.507	0.258	3.45E-19	3.30[2.49-4.39]	0.626	0.511	0.295	1.19E-25	2.56[2.12-3.08]
rs9266490	6	*HLA-B*	G	0.076	0.512	0.258	7.83E-19	3.15[2.40-4.14]	0.049	0.520	0.291	2.28E-26	2.52[2.11-3.02]

rs6910516	6	*HLA-B*	G	0.422	0.491	0.258	1.87E-15	2.85[2.17-3.75]	0.001	0.513	0.280	3.46E-25	2.51[2.09-3.02]
rs6937876	6	*PRDM1*	G	0.658	0.329	0.273	4.62E-02	1.29[1.00-1.65]	0.943	0.310	0.306	6.82E-01	
rs2285430	7	*HDAC9*	T	0.930	0.464	0.392	1.60E-02	1.34[1.05-1.70]	0.212	0.550	0.547	1.00	
rs6472266	8	*RRS1*	A	0.442	0.551	0.448	7.66E-04	1.51[1.19-1.93]	0.062	0.472	0.464	8.30E-01	
rs7832811	8	*SYBU*	G	0.566	0.283	0.225	2.18E-02	1.38[1.05-1.82]	0.070	0.272	0.255	3.87E-01	
rs888254	8	*CSMD3*	C	0.892	0.799	0.747	4.39E-02	1.32[1.01-1.72]	0.571	0.228	0.224	8.00E-01	
rs2149879	9	*SEMA4D*	G	1.000	0.339	0.252	9.69E-04	1.55[1.19-2.02]	0.952	0.304	0.289	5.07E-01	
rs303573	9	*CYLC2*	C	0.119	0.196	0.137	9.97E-03	1.49[1.10-2.03]	0.340	0.173	0.167	7.49E-01	
rs4749421	10	*LOC387647*	T	0.380	0.326	0.245	2.80E-03	1.48[1.14-1.93]	0.965	0.287	0.283	9.39E-01	
rs4749423	10	*LOC387647*	G	0.496	0.330	0.241	8.78E-04	1.55[1.20-2.02]	0.774	0.285	0.281	9.44E-01	
rs10824019	10	*PRKG1*	T	0.952	0.817	0.742	1.98E-03	1.59[1.18-2.13]	0.227	0.794	0.794	8.21E-01	
rs878824	10	*SH2D4B*	A	0.071	0.914	0.839	2.06E-04	1.92[1.35-2.78]	0.565	0.094	0.094	8.43E-01	
rs10786172	10	*CYP2C19*	G	0.366	0.311	0.222	9.11E-04	1.55[1.19-2.01]	0.756	0.292	0.260	6.67E-02	
rs7936044	11	*NELL1*	C	0.232	0.619	0.519	9.44E-04	1.47[1.16-1.85]	0.239	0.580	0.578	7.27E-01	
rs497279	11	*NOX4*	C	0.223	0.343	0.250	6.66E-04	1.56[1.20-2.02]	0.346	0.725	0.704	2.89E-01	
rs552221	11	*NTM*	C	0.607	0.218	0.142	1.09E-03	1.65[1.22-2.25]	0.902	0.807	0.803	7.01E-01	
rs17494112	12	*YAF2*	G	0.415	0.240	0.162	5.63E-04	1.71[1.26-2.34]	0.490	0.842	0.838	8.51E-01	
rs3852594	12	*GXYLT1*	G	0.835	0.240	0.161	7.67E-04	1.68[1.24-2.28]	0.275	0.845	0.835	5.19E-01	
rs6538225	12	*NAV3*	T	0.010	0.344	0.248	7.18E-04	1.53[1.19-1.95]	0.052	0.700	0.688	5.04E-01	
rs7206095	16	*SHISA9*	A	0.156	0.174	0.119	1.26E-02	1.51[1.09-2.08]	0.647	0.158	0.148	4.29E-01	
rs632111	19	*FUT2*	G	0.007	0.537	0.432	8.80E-04	1.46[1.17-1.82]	0.163	0.556	0.492	8.00E-03	1.23[1.05-1.45]
rs16983549	19	*PPP2R1A*	C	0.820	0.150	0.071	3.36E-05	2.20[1.49-3.25]	0.084	0.887	0.882	7.43E-01	
rs10417820	19	*PPP2R1A*	G	0.392	0.129	0.055	1.67E-05	2.52[1.62-3.91]	0.016	0.889	0.880	6.02E-01	
rs7256125	19	*PPP2R1A*	G	0.739	0.157	0.088	4.59E-04	1.88[1.31-2.71]	0.019	0.872	0.868	6.73E-01	
rs685119	20	*FAM110A*	C	0.270	0.562	0.498	4.00E-02	1.27[1.01-1.59]	0.296	0.564	0.544	3.89E-01	
rs7290471	22	*SAPS2*	G	1.000	0.534	0.457	1.18E-02	1.35[1.07-1.72]	0.916	0.485	0.476	7.17E-01	
rs2854131	MT	*MT-RNR2*	C	-	0.042	0.007	1.00E-04	6.18[2.13-17.91]	-	0.037	0.027	1.89E-01	

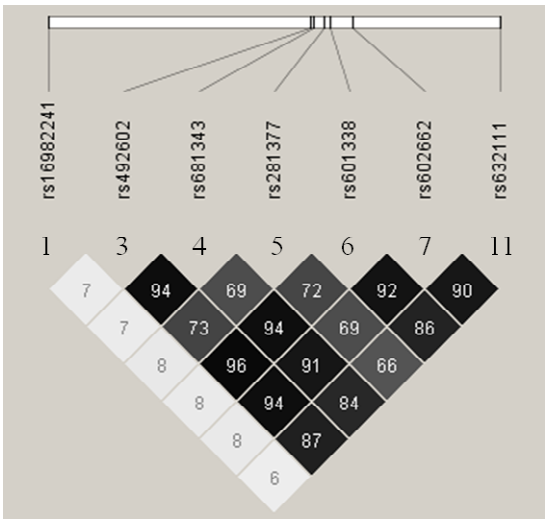
Supplementary Table 4. **Conditional regression analysis between *FUT2* associated SNPs in the Iranian combined dataset.** Pairwise *P*-values from the log-additive model regression analysis are shown for all SNPs. Significant *P*-values are highlighted in bold.

	Variables					
dbSNP ID	rs492602	rs681343	rs281377	rs601338	rs602662	rs632111
rs492602	-	3.23E-01	9.97E-02	4.32E-01	1.60E-01	7.49E-01
rs681343	5.68E-02	-	**1.33E-02**	3.37E-01	6.91E-01	1.20E-01
rs281377	6.21E-01	9.48E-01	-	9.36E-01	9.55E-01	5.42E-01
rs601338	1.28E-01	9.90E-01	**3.14E-02**	-	8.42E-01	2.80E-01
rs602662	**1.49E-02**	3.19E-01	**1.09E-02**	1.58E-01	-	9.48E-02
rs632111	2.17E-01	7.86E-01	5.26E-02	5.48E-01	8.14E-01	-

A.



B.



Supplementary Figure 1. Pairwise linkage disequilibrium (LD) plot for *FUT2* SNPs. Pairwise statistic r^2 values with white-to-black gradient shading proportional to the magnitude of LD ($r^2 = 0$: white; $0 < r^2 < 1$: shades of grey; $r^2 = 1$: black) were obtained in Haploview. A. This plot depicts the LD among *FUT2* SNPs with MAF>0.05 in 30 CEU HapMap Trios. B. This plot shows the LD among *FUT2* SNPs genotyped in the Iranian controls (N=826).

CHAPTER 7

General Discussion

7.1. MAIN FINDINGS

Behçet's disease is a complex disorder, and therefore, by contrast with single-gene disorders, is caused by many genetic and environmental factors working together, each having a relatively small effect and few if any being absolutely required for disease to occur [Manolio, 2010]. When this PhD began in January 2009, little was known about the genetics of Behçet's disease and *HLA-B*51* was the only established genetic risk factor for BD; however, its contribution to the overall genetic susceptibility to BD was estimated to be only 19%, or 32-52%, according to different studies [Gül *et al*., 2001; de Menthon *et al*., 2009]. Therefore, a substantial proportion of the BD genetic component was not yet accounted for. Furthermore, most of the studies had been performed in small samples sizes with association findings not replicated in independent samples.

The aim of this thesis was to contribute towards a better understanding of the genetic risk factors underlying Behçet's disease susceptibility by applying different independent approaches, such as the study of the mitochondrial genome, the follow up of known associations with BD, gene expression studies and a whole genome association study. These approaches allowed us to reach different findings that are discussed in this section.

7.1.1 Association of the mitochondrial polymorphism m.709G>A with Behçet's disease

Chapter 3 describes the association of the mitochondrial polymorphism 709G>A with Behçet's disease. m.709G>A is a non-coding variant located in the 12S rRNA gene (MT-RNR1), an RNA involved in the assembly of aminoacids into functional proteins [Ballana *et al*., 2006]. Mitochondrias have their own translation system for production of 13 proteins essential for oxidative phosphorylation in mammals and the mitochondrial ribosome is a central player in this translation system [Attardi *et al*., 1988]. To fulfil their main task, mitochondrias utilize reduction equivalents in form of NADH and molecular oxygen to generate an electrochemical potential across the inner mitochondrial membrane that is in turn utilized to generate ATP. Deficiencies in oxidative phosphorylation (OxPhoS) activity ultimately lead to a decrease in oxidative energy supply. Moreover, OxPhoS defects are accompanied by an increase of reactive oxygen species [Seibel *et al*., 2008]. There are several reports of highly pathogenic mtDNA mutations linked to mitochondrial diseases [Wallace *et al*., 1988; Guan, 2011], however the role of common mitochondrial variants in complex

diseases like the case of autoimmune diseases have been poorly investigated. Unlike pathogenic mutations, polymorphisms in the mitochondrial genome may not be the primary cause of disease but can contribute to modulate the risk and severity of immune mediated diseases, like BD, by increasing the oxidative stress and potentiating the inflammatory response (Figure 1). Although the m.709G>A does not appear to affect the secondary structure of the RNA, it may alter the tertiary or quaternary structure of this rRNA, or affect its decay rate [Xavier *et al*., 2011]. It is also possible that this variant is in LD with a coding variant that is itself the causative variant. Since in this study the mitochondrial variants tested for association with BD were select based on their ability to define the most common mitochondrial haplogroups in the Iranian population, a deeper analysis like sequencing of the entire 12S rRNA gene would be necessary to check for the existence of other variants in this gene with a stronger association with BD. This association finding indicated, for the first time, that in addition to multiple nuclear genes and environmental contributions, BD risk might also be governed by the mitochondrial genome. Therefore, this study warrants further validation in independent populations. The confirmation of the involvement of the m.709G>A with BD and of its potential effect in the increase of the oxidative stress can lead to advances in treatment options since it supports the hypothesis that the strengthening of anti-oxidant defenses may constitute an additional and complementary treatment strategy in these patients.

7.1.2 Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients

Chapter 4 describes the replication of the association of common variants at the *IL10* and *IL23R-IL12RB2* locus, previously identified as BD risk factors in two GWAS [Remmers *et al.,* 2010; Mizuki *et al.,* 2010]. We identified two variants in *IL10* nominally associated and several variants in *IL23R-IL12RB2* strongly associated with BD, in the Iranian population [Xavier *et al*., 2012]. Our SNP imputation analysis, followed by genotyping validation, allowed the identification of the regulatory region upstream of *IL23R* as the most strongly associated one. Both *IL10* and *IL23R* have been previously associated with several autoimmune diseases [Franke *et al*., 2008; Gateva *et al*., 2009; Safrany *et al*., 2009] and, interestingly, both IL10 and IL23 cytokines have been pointed out as important players in the immunological response observed in Behçet's disease patients (as referred in chapter 1.1.8.4 and 1.8.8.5, pages 13 to 15).

IL10 is associated with many autoimmune diseases because of its anti-inflammatory functions [Asadullah *et al.*, 2003]. In fact, IL10 is an important cytokine in the induction of Th2 response that plays a crucial role in adaptive immunity via the induction of specific antibodies to eliminate the reinvasion of microbes and the absorption of microbial products. IL10 is also one of the most effective immune regulatory cytokines in contributing to maintaining the homeostasis of the body, protecting against the epithelial damage caused by inflammation [Barnes *et al.*, 2009; Jarry *et al.*, 2008]. Although $Il10^{-/-}$ mice develop normally and have apparent normal antibody responses, these mice spontaneously develop colitis that is characterized by abnormal architecture in the mucosa associated with the infiltration of various leukocyte subsets [Kuhn *et al.*, 1993]. IL10 inhibits directly macrophages and DCs from producing IL12, a key driving factor for Th1 differentiation and also inhibits the subunits p40 and p35 of IL12, at a transcriptional level in myeloid cells [Aste-Amezaga *et al.*, 1998]. As the p40 subunit is shared by IL23 [Oppmann *et al.*, 2000], IL-10 can inhibit IL-23 production from DCs and macrophages. Therefore, IL-10 is an essential negative regulator, especially for controlling IL-23 induction, during mucosal inflammation, that is a feature of Behçet's disease. Studies have shown a suppression of the production of IL10 in the intestine of inflammatory bowel disease patients [Fuss *et al.*, 1996] and that, as expected given its anti-inflammatory properties, the IL10 levels correlated negatively with the infiltration of inflammatory cells in the colon mucosa of this patients [Liu *et al.*, 2012]. Interestingly it was also observed that IL23 played a critical role in the suppression of IL10 transcription in polarized CD4+ T cells [Liu *et al.*, 2012].

IL23 is required for the maintenance of Th17 cells, inducing the production of the lineage specific effector cytokines. The acute response to danger signals received in the skin, or gastrointestinal tract, includes immediate activation of the IL23-IL17 pathway promoting pathogen clearance, inadequately controlled by Th1 or Th2 cells. This result in the recruitment of neutrophils and activation of macrophages to the infectious site [Elson *et al.*, 2007], and in the production of oxygen radicals and other inflammatory mediators, thereby, providing immediate protection against the microbial invasion. Once the insult has been identified, appropriate antigen-specific effector mechanisms will replace the granulocyte response [McKenzie *et al.*, 2006]. However the cost of having this potent cell-mediated immunity is very high. If the IL23-IL17 immune pathway becomes deregulated, there is a danger of breaking tolerance to self tissues and antigens, leading to severe autoimmune pathologies [McKenzie *et al.*, 2006] and, in fact, these cells have been showed to have a dominant role in provoking chronic autoimmune inflammation of the central nervous

system and joints [Languish *et al.*, 2005; Cue *et al.*, 2003; Murphy *et al.*, 2003]. Consistent with these observations, transgenic mice ubiquitously over expressing IL23 develop severe multi-organ inflammation with elevated expression of pro-inflammatory cytokines such as TNF and IL-1. In addition to driving Th17 development, thereby promoting chronic inflammatory responses, IL23 functions may also be closely related to the influx of neutrophils during acute infections [Chen *et al.*, 2006], a feature observed in BD histopathology lesions.

Therefore, an imbalance between the anti-inflammatory and pro-inflammatory properties of IL10 and IL23, respectively, seem to modulate Behçet's disease pathogenic process, influencing the development of the disease and maybe modifying its course.

Narrowing an implicated locus to the single variant that actually causes susceptibility to disease by disrupting the expression or function of a protein has proven difficult in practice [Manolio, 2010]. But Remmers and colleagues showed that rs1518111, the *IL10* promoter variant more associated with BD in the original GWAS and in our replication analysis, was associated with a diminished mRNA expression and low protein production and therefore the anti-inflammatory properties of IL10 may be compromised in BD patients [Remmers *et al.*, 2010]. Our study highlighted the importance of replicating association findings across different populations, in order to narrow the associated regions and identify SNPs associated in multiple populations, that are likely to be in LD with the causal variant [Stranger *et al.*, 2011]. The definitive validation of these associations is of foremost importance as it may have therapeutic implications, namely the specific targeting of the IL23/IL17 axis that has been observed to be impaired in several auto-immune disorders and more recently in BD.

7.1.3 Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility

Chapter 5 describes the implication of the neuregulin signaling pathway in Behçet's disease through both gene expression and association studies. We combined gene expression with *in silico* pathway analysis in order to select candidate genes to be tested for association with BD, without using *a priori* assumptions on gene function and possible disease mechanisms. Association analysis of *EREG*, *AREG* and *NRG1,* which were down-regulated in BD patients, allowed the identification of a novel association at a variant located downstream to *EREG* and of three variants in *NRG1* [Xavier *et al.*, 2013]. EREG, AREG and NRG1 are members of the epidermal growth factor (EGF) family and have highly similar structural and

functional characteristics. These association findings support a role for the EGF family and receptors signaling pathway in BD pathogenesis, that warrants further investigation.

The activity of EGF family members is mediated by the epidermal growth factor receptor tyrosine kinases (EGFR/ErbB), which are expressed on most human cell types [Pastore *et al*., 2008]. Receptor-ligand interactions induces the heterodimerization of receptor monomers, which in turn results in the activation of intracellular signaling cascades and the induction of cellular responses including proliferation, migration, differentiation, and survival or apoptosis [Britsch., 2007]. *EREG*, that was the most underexpressed gene in our study, is already known to play an essential role in immune/inflammatory-related responses in the epidermal layer [Shirasawa *et al*., 2004] and has a pivotal role in cytokine production by antigen presenting cells; *AREG* seems to act like a possible mediator of innate cutaneous immunity and epidermal proliferation, and a potential trigger of inflammatory reaction [Cook *et al*., 2004]; and *NRG1* has been also linked to a possible immune deregulation [Marballi *et al*., 2010]. The observation that epithelial cancers are characterized by overexpression of epidermal growth factor receptors (EGFR) has led to the development of therapeutic strategies that block these growth factor receptors [Pastore *et al*., 2008]. Interesting common adverse affects of this therapy include pustular or acneiform eruptions (which can be severe enough to lead to treatment modification or cessation), characterized histopathologically by a moderate to severe inflammatory reaction dominated by neutrophils which surround and then invade follicular infundibula [reviewed in Pastore *et al*., 2008], a phenotype that resembles BD skin lesions. In the mouse models of irritant contact dermatitis and allergic contact dermatitis, EGFR inhibition leads to aggravation of the skin inflammatory response, with upregulated chemokine expression and massive skin infiltration by T cells and macrophages [reviewed in Pastore *et al*., 2008]. These data strongly suggest that pharmacological abrogation of EGF signaling pathway typically exerts peculiar inflammatory/toxic effects on the skin by impairing the multiple EGFR-dependent homeostatic functions of the skin [Pastore *et al*., 2008]. Our results suggest that under-expression of this pathway is associated to Behçet's disease pathogenesis, and that the inflammatory reaction observed in BD patients may be mediated or modified by this signaling pathway. To further establish the role of this pathway in BD pathogenesis, comprehensive immunogenetic/cellular studies may be conducted.

Epistasis analyses also detected the presence of interactions between *EREG* and *NRG1* variants, meaning that the joint effect of alleles in these genes is higher than the sum of the effect anticipated for each allele independently. Therefore, the marginal effect of each gene

may be small, however by acting together, epistatically, in the Neuregulin signaling pathway, this interaction may reflect much larger effects. This epistatic measure is dependent on the ability of this combination of SNPs to correctly classify the samples (cases or controls) and, in general, does not have a clear interpretation in terms of mechanism of interaction in the biological network [Cordell *et al*., 2005]. This study also indicates that *EREG-AREG* and *NRG1* seem to modulate BD susceptibility through main effects and also gene-gene interactions. It also highlights the importance of combining genetic and genomic approaches to dissect the genetic architecture of complex diseases.

7.1.4 *FUT2*: filling the gap between genes and environment in Behçet's disease?

Chapter 6 describes the first genome-wide association study conducted in the Iranian population. In this study, we used a DNA pooling approach and identified the *FUT2* as a novel genetic risk factor for BD. The missense variant in *FUT2* most strongly associated with BD in our Iranian dataset (rs602662) is located in a haplotype block with other functional variants, such as rs601338 (W143X) that codifies for a stop codon and was also associated in our dataset. Although, rs601338 is responsible for the non-secretor phenotype in a large proportion of the Caucasian population, we cannot exclude the possibility that rs602262, together with rs601338, contribute to the non-secretor phenotype in the Iranian population. The non-secretor phenotype has been observed to alter the gut microbiome [Wacklin *et al*., 2011] and to be associated with resistance to a variety of infectious diseases such as Norovirus [Carlsson *et al*., 2009] and Helicobacter pylori [Ikehara *et al*., 2001]. More studies will be necessary to better understand how the secretor phenotype encoded by *FUT2* modulates BD risk: if by causing changes in the mucosa glycosylation and adherence, and protecting the host against several types of infections, having as a consequence a decrease in the antigenic stimulation in the early life of the individuals, and consequent increased propensity for auto-immunity [Yang *et al*., 2011]; if by changing the resistance to the trigger factor that initiates the cascade of immunological events present in BD pathogenesis; or by any other mechanism. This novel *FUT2* association is of great scientific interest since it may constitute the missing link between the genetic and the environmental risk factors in BD susceptibility.

The variants located in the *HLA-B* locus were, as expected, the most strongly associated with BD. In fact, *HLA-B*51* constitutes the most accurate genetic marker for BD to date in different ethnic groups. Therefore we did not explore further (e.g., fine-mapping,

allele typing) the association of this locus with BD in this particular study, nor in the remaining work reported in this thesis.

We also identified an association at rs7528842, located in a gene desert, closest to the Diphthamide biosynthesis methyltransferase (*DPH5)* gene. The association of this variant with BD was replicated in an independent Iranian dataset but not in the Turkish, and therefore was not further explored. GWASs routinely implicate variation within gene deserts and other types of noncoding DNA. GWASs are based on the premise that if the causal variant is located on a haplotype, then the genotyping of a marker allele in LD with the causal variant will likely show an association with the phenotype of interest. However, in 40% of cases, known exonic sequences are absent from the associated LD blocks [Visel *et al*., 2009]. While the presence of non-annotated transcripts or non-coding RNAs may explain some of these association hits, many of the associated non-coding regions may also harbor variants that alter the activity of long-range *cis*-regulatory elements controlling gene expression [Wasserman *et al*., 2010]. Despite the abundance of GWAS signals implicating noncoding regions in complex disease risk, strategies to experimentally follow up on such findings are lacking [Wasserman, *et al*., 2010].

7.2 LIMITATIONS OF THE EXPERIMENTAL DESIGN

7.2.1 Dataset

An important topic in genetic association studies is the sample size which should have adequate statistical power in order to achieve robust and reproducible results [Collins, 2009; Palmer *et al*., 2005]. During the development of this thesis we created a high-quality and well-characterized biobank, including about 1000 BD patients and 800 controls from Iran. For a relatively rare disease such as BD, a dataset of this dimension has a good power to detect common variants conferring risk to disease. Also, our BD cases represents the second largest dataset used so far in published genetic association studies for BD, being only smaller than a Turkish dataset used in previous GWAS [Remmers *et al*., 2010; Kirino *et al*., 2012].

Our case/control Iranian dataset was collected between 2007 and 2012 and therefore the first studies performed in the context of this thesis did not include the entire dataset and were performed in a smaller number of samples. Furthermore, the group of controls included in the first study performed (Chapter 3 - Association of mitochondrial polymorphism m.709G>A with Behçet's disease) included controls with recurrent oral aphthae, that were excluded from posterior studies, as those shown in Chapter 4 to 6. The rationale behind the inclusion of the controls with ROA in the first study presented in this thesis was that these samples were representative of the general population (where the ROA also occur at a frequency of 20%) [Porter *et al*., 1998]. Even if this group of controls was enriched in genetic risk factor for BD, when compared to controls without ROA, this would not lead to an increase in the type I error (false positive findings). However, after the publication of this study, we decided to exclude the controls with ROA from further studies since the inclusion of these samples could lead to a decrease in power and to missing true association findings and to include more 400-500 controls without OA.

7.2.2 Diagnostic criteria

Regarding the issue of appropriate phenotype characterization in genetic association studies, the BD samples used in this thesis were selected according to the revised International criteria for Behçet's disease (ICBD) [ITR-ICBD, 2006]. However, almost half of these BD cases also fulfill the more specific International Study Group Criteria for Behçet's Disease (ISG) [ISG, 1990]. In Chapter 3 of this thesis - Association of mitochondrial

polymorphism m.709G>A with Behçet's disease - we tested the association of the mitochondrial genome in both datasets of cases. We found that the variant m.709G>A had a stronger association in the ISG group of cases than in the ICBD group, leading us to propose that the ISG group might be a more genetically homogenous group of cases [Xavier *et al*., 2010]. One question must be raised with this finding: is it better to have a bigger but possibly less homogenous group of cases or a more homogeneous group of cases at the cost of reduced sample size? Since no general answer is possible and all the cases were diagnosed according to published international criteria and by physicians specialists in the field, we decided to include the entire dataset of cases, fulfilling the ICBD, in the subsequent studies.

7.2.3 Adjustment for confounding variables

As mentioned in chapter 1.2.2.4 (pages 40 to 43), an important point in the analysis of genetic associations is the adjustment for potentially confounding variables. Behçet's disease has no established environmental or epidemiological risk factors. Several auto-immune diseases are more frequent in females than males, like rheumatoid arthritis and SLE; however, in BD, there is no obvious gender imbalance since in eastern Mediterranean populations the disease is more common in men, but in Asian populations the sex ratio is reversed [Marshall, 2004]. Yet, in three of the studies reported in this thesis there was an imbalance in the male to female ratio between cases and controls, and the question of whether to adjust the association analysis for gender was debated. We decided to perform both unadjusted and adjusted analyses for gender, whenever the gender ratio was imbalanced between cases and controls, in order to have a more conservative approach regarding the possible effect of gender in the risk conferred by the variant tested for association. In Chapter 3 - Association of mitochondrial polymorphism m.709G>A with Behçet's disease - we showed both adjusted and unadjusted association results, which were very similar; in Chapter 4 - Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients - there was no difference between cases and controls in the gender distribution and, for that reason, no adjustment for gender was performed; in Chapter 5 - Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility – the male to female ratio was significantly increased in the BD group but, since no major differences were observed between the adjusted and unadjusted analyses, the results were reported unadjusted; and finally in Chapter 6 - *FUT2*: filling the gap between genes and environment in Behçet's disease? - the

proportion of males was also higher in the disease group when compared to the control group (in the replication dataset and in the combined dataset) and although the results adjusted and unadjusted were very similar, this time we reported the results adjusted for gender. Besides the *HLA-B*51/B*5* carriage which seems to predominate in male BD patients [Maldini *et al*., 2012], there is no evidence suggesting the existence of other gender-biased genetic risk factors for BD. The results shown in this thesis further support this evidence, since similar results were obtained before and after adjustment for gender, both for the mitochondrial and nuclear genome. Therefore, we believe that when we are testing a new variant for association with BD, in a sample of cases and controls with differences in the gender frequency, association tests should be performed unadjusted and adjusted for gender although, much likely, they will not show significant differences. If this is the case, non-adjusted results may me reported, or they may be adjusted if we want to follow a more conservative approach.

Stratification is another confounding variable that is frequently not well considered in association studies and that can result in spurious association findings. In the first study performed in the context of this thesis, described in Chapter 3 - Association of mitochondrial polymorphism m.709G>A with Behçet's disease - population stratification was assessed using a panel of 89 autosomal ancestry informative markers, in a dataset of 434 controls and 615 BD cases, and no evidence of stratification was observed [Xavier *et al*., 2011]. However, in the study presented in Chapter 4 - Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients - the sample used was larger, composed by 637 controls and 973 BD patients and therefore we decided to test again for the existence of population stratification, using the same ancestry informative markers. In this study, the ANOVA statistics for population differences along the eigenvectors, demonstrated that BD cases and controls did not differ on the first two principal components and in the overall test and, therefore, the association findings were not likely to be confounded by hidden population stratification. However, since there was a significant effect of affection status on some of the eigenvectors, and this could indicate the presence of subtle ancestry differences, we decided to use the first 13 principal components as covariates for ancestry adjustment in the association tests [Xavier *et al*., 2012]. It is noteworthy that the corrected and uncorrected association results were very similar and therefore we concluded that in future studies, like those shown in Chapter 5 and 6, it would be not necessary to correct for population stratification.

7.2.4 htSNPs approach

Due to the existence of LD blocks in the genome, genotyping of all the variants within a region can become redundant. Thus, we adopted a haplotype tagging SNP genotyping approach in several studies reported in this thesis. Tagging exploits the extensive linkage disequilibrium in many parts of the genome. Therefore, by typing the subset of variants that capture almost all of the genetic diversity existing in a given genomic region, it is possible to maintain power while making considerable savings in genotyping [Hattersley *et al*., 2005]. We have used the HapMap CEU population as reference for both selecting htSNPs to be genotyped and also to perform imputation analyses. Before we performed these studies, there was no published data about selection of htSNPs or genotype imputation analysis in the Iranian population and consequently there was a concern about whether the LD patterns existing in the CEU HapMap population were representative of the LD patterns occurring in the Iranian population. We have performed this comparison in Chapter 4 - Association study of *IL10* and *IL23R-IL12RB2* in Iranian Behçet's disease patients – and our results strongly supports the use of this HapMap dataset for selecting tagging SNPs and as a reference to perform genotype imputation in the Iranian population [Xavier *et al*., 2012].

7.2.5 OR observed

Generally associations between SNPs and complex traits tend to be of modest effect size, with a median odds ratio per copy of the risk allele of 1.33 [Hindorff *et al*., 2009] and thus, poor predictors of disease development. However if they are common variants, as the case of the variants described as associated with BD in this thesis, they may collectively account for a large portion of the disease risk experienced by the affected population [Juran *et al*., 2007]. However, for most of the complex diseases including Behçet's diseases, there is still a missing heritability, including much larger numbers of variants of smaller effect yet to be found, rarer variants (possibly with larger effects) that are poorly detected by available genotyping arrays that focus on common variants, structural variants poorly captured by existing arrays and low power to detect gene-gene interactions [Manolio *et al*., 2009]. Therefore, the genetic basis underlying BD should continue to be investigated, but new approaches and new methods of analysis should be applied in order to identify new and other types of risk variants.

7.3 FOLLOW-UP STUDIES

The most comprehensive approach to characterize the contribution of low frequency and rare variants that are likely to have larger effects on complex traits than common variants is through large-scale, next-generation re-sequencing studies. Despite advances in the cost-effectiveness of these technologies, whole-genome or even exome re-sequencing of the large cohorts of individuals required to detect rare variant association with complex traits, still represent an infeasible financial undertaking for most research groups. However, with the advent of the 1000 Genome (1KG) project, high-density reference panels obtained from whole-genome re-sequencing data are being released. The 1KG provided a comprehensive catalogue of variation with MAF as low as 0.5%, as well as many rarer variants, across a wide range of populations from different ethnic groups [1000 Genomes Project Consortium *et al*., 2010]. Such reference panels can be utilized to select rare variants for large-scale genotyping with custom designed arrays, potentially with priority given to variants with likely functional consequences in an effort to reduce costs. If, in addition, samples have already been assayed by GWAS or by candidate gene association analysis, imputation techniques [Marchini *et al*., 2010] can make use of this existing scaffold to predict genotypes at variants present in the higher density reference panel, incurring no additional cost, other than computation. This technique could be applied for genes densely genotyped in this thesis, as *IL10*, *IL23R-IL12RB2* or *FUT2* genes, in order to identify rare missensse variants present in our case-control dataset that could be posteriorly validated by genotyping of the samples, and tested for association with BD. However, the effects of rare variants are unlikely to be sufficiently large to be detected through association testing of individual rare variants. Therefore, statistical methods that focus on the aggregation of the effects of all lower frequently variants within the same exon, gene or pathway, potentially weighting according to annotation or MAF, can be applied. Using these methods, multiple rare variants have been demonstrated to be associated with a variety of complex traits including low- and high-density lipoprotein [Cohen *et al*., 2007; Romeo *et al*., 2006], body mass index [Ahituv *et al*., 2006] and blood pressure [Ji *et al*., 2008].

7.4 FINAL REMARKS

The vast majority of human diseases are genetically complex, wherein the direct correspondence between causative genotype and disease phenotype characteristic of Mendelian disorders is not present [Strohman, 2002]. Instead, complex diseases develop as the cumulative result of environmental exposures, exerting their effect over time, in genetically susceptible individuals. Therefore, the genotypic components of complex diseases are not causative but rather mediate disease risk [Juran *et al*., 2007]. Many such susceptibility genotypes are expected for each complex disorder, some of which are common to similar diseases (e.g., *IL23R* in autoimmune disorders) and some of which may be disease specific (e.g., *HLA-B*51* in BD). Some of them may also establish a link between disease and environment, like the new finding of the association of *FUT2* wit BD reported in this thesis. Regardless, each individual variant is likely to have only a slight contribution to the overall risk of each specific disease [Juran *et al*., 2007].

The ultimate goal in medical genetics is the identification of causal functional variants, or variants that increase risk for disease, in order to elucidate the mechanisms through which they exert their effects and, ultimately, result in more personalized and effective prevention strategies and therapeutic options. Therefore, trait mapping studies can be considered hypothesis-generating exercises, helping to prioritize genes or genomic regions for further investigation. I believe that during the work developed under this PhD thesis we have accomplished this goal by pointing out the mitochondrial genome, the neuregulin signaling and the *FUT2* gene as novel players in genetic BD susceptibility that should be followed-up since they may contribute to the abnormal immunological response observed in Behçet's disease (Figure 1). We have also contributed to establish the *IL10* and *IL23R* loci as worldwide risk factors for Behçet's disease, cytokines that ate known to play an important role in the immune response, and whose deregulation may also be responsible for some pathological features observed in BD patients (Figure 1)

It is important to mention though, that variants in both *NRG1* and *FUT2* have been previously found associated in BD GWAS [Remmers *et al*., 2010; Mizuki *et al*., 2010; Kirino *et al*. 2013]. However, as these variants did not reach the genome-wide significance thereshold applied in these studies, they were not reported in the main manuscripts. The studies performed in the context of this thesis allowed to highlight the association of BD with these genes and to point out the importance of combining findings genome-wide below in order to identify disease risk genes consistent across different population.

Narrowing an implicated locus to a single variant that directly causes susceptibility to disease will be a key step in improving our understanding of the mechanisms of disease and in designing effective strategies for risk assessment and treatment [Manolio, 2010].

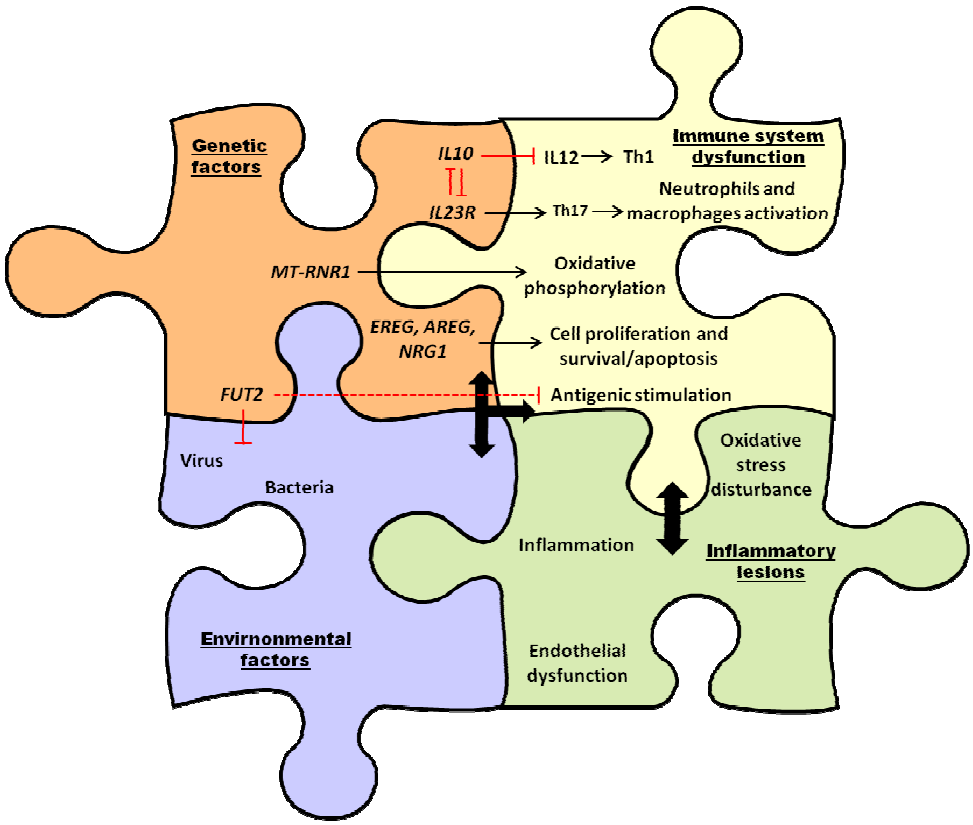


Figure 1. Diagram presenting the putative pathogenic mechanisms of the genes associated with BD in this thesis. The genes reported as associated with Behçet's disease in this thesis – *MT-RNR1, IL10, IL23R, EREG-AREG, NRG1* and *FUT2* – are shown, together with their possible mechanisms of action in the immune system dysfunction and inflammatory lesions, in the context of the current model for Behçet's disease pathogenesis.

7.5 REFERENCES

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Collier JM, Hébert S, Doelle H, Dent R, Pennacchio LA, McPherson R. A PYY Q62P variant linked to human obesity. Hum Mol Genet. 2006;15(3):387-91.

Asadullah K, Sterry W, Volk HD. Interleukin-10 therapy—review of a new approach. Pharmacol Rev. 2003;55(2):241-69.

Aste-Amezaga M, Ma X, Sartori A, Trinchieri G. Molecular mechanisms of the induction of IL-12 and its inhibition by IL-10. J Immunol. 1998;160:5936–44.

Attardi G, Schatz G. Biogenesis of mitochondria. Annu Rev Cell Biol. 1988;4:289-333.

Ballana E, Morales E, Rabionet R, et al., Mitochondrial 12S rRNA gene mutations affect RNA secondary structure and lead to variable penetrance in hearing impairment Biochem Biophys Res Commun 2006;341:950-7.

Barnes MJ, Powrie F. Regulatory T cells reinforce intestinal homeostasis. Immunity. 2009;31(3):401-11.

Britsch S.The neuregulin-I/ErbB signaling system in development and disease. Adv Anat Embryol Cell Biol. 2007;190:1-65. Review.

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature 2004; 429: 446–52.

Carlsson B, Kindberg E, Buesa J, Rydell GE, Lidón MF, Montava R, Abu Mallouh R, Grahn A, Rodríguez-Díaz J, Bellido J, Arnedo A, Larson G, Svensson L. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. PLoS One. 2009;4(5):e5593.

Chen Y, Languish CL, McKenzie B, Joyce-Shaikh B, Stumhofer JS, McClanahan T, Blumenschein W, Churakovsa T, Low J, Presta L, Hunter CA, Kastelein RA, Cue DJ. Anti-IL-23 therapy inhibits multiple inflammatory pathways and ameliorates autoimmune encephalomyelitis. J Clin Invest. 2006;116(5):1317-26.

Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A. 2006;103(6):1810-5.

Collins A. Approaches to the identification of susceptibility genes. Parasite Immunol. 2009;31(5):225-33.

Cook PW, Brown JR, Cornell KA, Pittelkow MR. Suprabasal expression of human amphiregulin in the epidermis of transgenic mice induces a severe, early-onset, psoriasis-like skin pathology: expression of amphiregulin in the basal epidermis is also associated with synovitis. Exp Dermatol. 2004;13(6): 347-56.

Cordell HJ, Clayton DG. Genetic association studies. Lancet. 2005;366(9491): 1121-31.

Cue DJ, Sherlock J, Chen Y, Murphy CA, Joyce B, Seymour B, Lucian L, To W, Kwan S, Churakova T, Zurawski S, Wiekowski M, Lira SA, Gorman D, Kastelein RA, Sedgwick JD Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature. 2003;421(6924):744-8.

de Menthon M, Lavalley MP, Maldini C, Guillevin L, Mahr A. HLA-B51/B5 and the risk of Behçet's disease: a systematic review and meta-analysis of case-control genetic association studies. Arthritis Rheum. 2009;61(10): 1287-96. Review.

Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sacli FS, Erer B, Inoko H, Emrence Z, Cakar A, Abaci N, Ustek D, Satorius C, Ueda A, Takeno M, Kim Y, Wood GM, Ombrello MJ, Meguro A, Gül A, Remmers EF, Kastner DL. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013;45(2):202-7.

Elson CO, Cong Y, Weaver CT, Schoeb TR, McClanahan TK, Fick RB, Kastelein RA. Monoclonal anti-interleukin 23 reverses active colitis in a T cell-mediated model in mice. Gastroenterology. 2007;132(7):2359-70.

ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306, 636–640 (2004).

Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH; IBSEN study group, Mathew CG, Schreiber S. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008;40:1319–23.

Fuss IJ, Neurath M, Boirivant M, Klein JS, de la Motte C, Strong SA, Fiocchi C, Strober W. Disparate CD4+ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn disease LP cells manifest increased secretion of IFN-γ, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. J. Immunol. 1996;157(3):1261-70.

Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, Seldin MF, Rönnblom L, Criswell LA, Syvänen AC, Behrens TW, Graham RR. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet 2009;41:1228–33.

Guan MX. Mitochondrial 12S rRNA mutations associated with aminoglycoside ototoxicity. Mitochondrion. 2011;11:237–45.

Gül A, Hajeer AH, Worthington J, Barrett JH, Ollier WE, Silman AJ. Evidence for linkage of the HLA-B locus in Behçet's disease, obtained using the transmission disequilibrium test. Arthritis Rheum. 2001;44(1):239-40.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362-7.

Ikehara Y, Nishihara S, Yasutomi H, Kitamura T, Matsuo K, Shimizu N, Inada K, Kodera Y, Yamamura Y, Narimatsu H, Hamajima N, Tatematsu M. Polymorphisms of two fucosyltransferase genes (Lewis and Secretor genes) involving type I Lewis antigens are associated with the presence of anti-Helicobacter pylori IgG antibody. Cancer Epidemiol Biomarkers Prev. 2001;10(9):971-7.

International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD). Revision of the International Criteria for Behçet's Disease (ICBD). Clin Exp Rheumatol 2006;24(suppl 42): S14-S5.

International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. Lancet. 1990;335(8697): 1078-80.

Jarry A, Bossard C, Bou-Hanna C, Masson D, Espaze E, Denis MG, Laboisse CL. Mucosal IL-10 and TGF-β play crucial roles in preventing LPS-driven, IFN-γ-mediated epithelial damage in human colon explants. J. Clin. Invest. 2008;118(3):1132-42.

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet. 2008;40(5):592-9.

Juran BD, Lazaridis KN. Applying genomics to the study of complex disease. Semin Liver Dis. 2007 Feb;27(1):3-12.

Kuhn R, Lohler J, Rennick D, Rajewsky K, Muller W. Interleukin-10-deficient mice develop chronic enterocolitis. Cell. 1993;75:263–74.

Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, McClanahan T, Kastelein RA, Cua DJ. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. J Exp Med. 2005;201(2):233-40.

Liu Z, Feng BS, Yang SB, Chen X, Su J, Yang PC. Interleukin (IL)-23 suppresses IL-10 in inflammatory bowel disease. J Biol Chem. 2012;287(5):3591-7.

Maldini C, Lavalley MP, Cheminant M, de Menthon M, Mahr A. Relationships of HLA-B51 or B5 genotype with Behcet's disease clinical characteristics: systematic review and meta-analyses of observational studies. Rheumatology (Oxford). 2012;51(5): 887-900. Review.

Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363(2):166-76.

Marballi K, Quinones MP, Jimenez F, Escamilla MA, Raventós H, Soto-Bernardini MC, Ahuja SS, Walss-Bass C. In vivo and in vitro genetic evidence of involvement of neuregulin 1 in immune system dysregulation. J Mol Med (Berl). 2010;88(11):1133-41.

Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511.

Marshall SE. Behçet's disease. Best Pract Res Clin Rheumatol. 2004; 18(3): 291-311. Review.

McKenzie BS, Kastelein RA, Cua DJ. Understanding the IL-23-IL-17 immune pathway. Trends Immunol. 2006;27(1):17-23.

Mizuki N, Meguro A, Ota M, Ohno S, Shiota T, Kawagoe T, Ito N, Kera J, Okada E, Yatsu K, Song YW, Lee EB, Kitaichi N, Namba K, Horie Y, Takeno M, Sugita S, Mochizuki M, Bahram S, Ishigatsubo Y, Inoko H. Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat Genet. 2010;42(8):703-6.

Murphy CA, Langrish CL, Chen Y, Blumenschein W, McClanahan T, Kastelein RA, Sedgwick JD, Cua DJ. Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. J Exp Med. 2003;198(12):1951-7.

Oppmann B, Lesley R, Blom B, Timans JC, Xu Y, Hunte B, Vega F, Yu N, Wang J, Singh K, Zonin F, Vaisberg E, Churakova T, Liu M, Gorman D, Wagner J, Zurawski S, Liu Y, Abrams JS, Moore KW, Rennick D, de Waal-Malefyt R, Hannum C, Bazan JF, Kastelein RA. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity. 2000;13(5):715-25.

Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet. 2005;366(9492): 1223-34.

Pastore S, Mascia F, Mariani V, Girolomoni G. The epidermal growth factor receptor system in skin repair and inflammation. J Invest Dermatol. 2008;128(6):1365-74.

Porter SR, Scully C, Pedersen A. Recurrent aphthous stomatitis. Crit Rev Oral Biol Med. 1998;9:306-21.

Remmers EF, Cosan F, Kirino Y, Ombrello MJ, Abaci N, Satorius C, Le JM, Yang B, Korman BD, Cakiris A, Aglar O, Emrence Z, Azakli H, Ustek D, Tugal-Tutkun I, Akman-Demir G, Chen W, Amos CI, Dizon MB, Kose AA, Azizlerli G, Erer B, Brand OJ, Kaklamani VG, Kaklamanis P, Ben-Chetrit E, Stanford M, Fortune F, Ghabra M, Ollier WE, Cho YH, Bang D, O'Shea J, Wallace GR, Gadina M, Kastner DL, Gül A. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet. 2010;42(8):698-702.

Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet. 2007;39(4):513-6.

Safrany E, Melegh B. Functional variants of the interleukin-23 receptor gene in non-gastrointestinal autoimmune diseases. Curr Med Chem 2009;16:3766–74.

Seibel P, Di Nunno C, Kukat C, Schäfer I, Del Bo R, Bordoni A, Comi GP, Schön A, Capuano F, Latorre D, Villani G. Cosegregation of novel mitochondrial 16S rRNA gene mutations with the age-associated T414G variant in human cybrids. Nucleic Acids Res. 2008;36(18):5872-81.

Shirasawa S, Sugiyama S, Baba I, Inokuchi J, Sekine S, Ogino K, Kawamura Y, Dohi T, Fujimoto M, Sasazuki T. Dermatitis due to epiregulin deficiency and a critical role of epiregulin in immune-related responses of keratinocyte and macrophage. Proc Natl Acad Sci U S A. 2004;101(38): 13921-6.

Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. 2011;187(2):367-83.

Strohman R. Maneuvering in the complex path from genotype to phenotype. Science. 2002;296(5568):701-3.

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009;461(7261):199-205.

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ 2nd, Nikoskelainen EK. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science. 1988;242(4884):1427-30.

Wacklin P, Mäkivuokko H, Alakulppi N, Nikkilä J, Tenkanen H, Räbinä J, Partanen J, Aranko K, Mättö J. Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. PLoS One. 2011;6(5):e20113.

Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. 2010;20(9):1191-7.

Xavier JM, Shafiee NM, Ghaderi F, Rosa A, Abdollahi BS, Nadji A, Shahram F, Davatchi F, Oliveira SA. Association of mitochondrial polymorphism m.709G>A with Behçet's disease. Ann Rheum Dis. 2011;70(8):1514-6.

Xavier JM, Shahram F, Davatchi F, Rosa A, Crespo J, Abdollahi BS, Nadji A, Jesus G, Barcelos F, Patto JV, Shafiee NM, Ghaderibarim F, Oliveira SA. Association study of IL10 and IL23R-IL12RB2 in Iranian patients with Behçet's disease. Arthritis Rheum. 2012;64(8):2761-72.

Xavier JM, Krug T, Davatchi F, Shahram F, Fonseca BV, Jesus G, Barcelos F, Vedes J, Salgado M, Abdollahi BS, Nadji A, Moraes-Fontes MF, Shafiee NM, Ghaderibarmi F, Vaz Patto J, Crespo J, Oliveira SA. Gene expression profiling and association studies implicate the neuregulin signaling pathway in Behçet's disease susceptibility J Mol Med (Berl). 2013 Apr 27. [Epub ahead of print]

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.